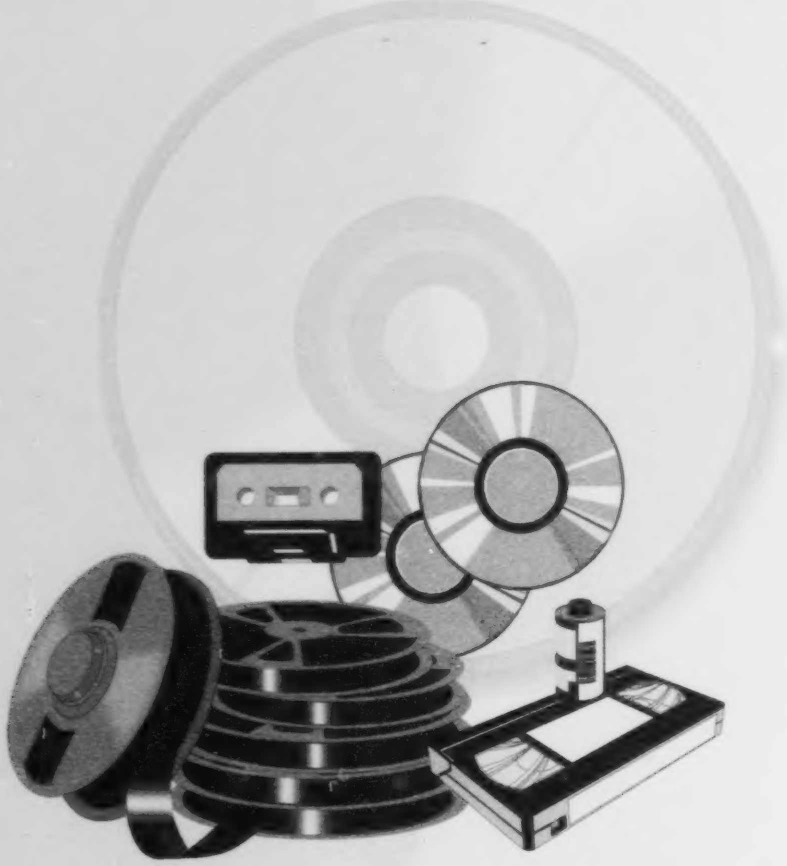
Canadian
Conservation
Institute

Institut
canadien de
conservation

Preservation of Electronic Records: New Knowledge and Decision-making

La préservation des documents électroniques : Information récente et prise de décisions



Canadian Heritage Patrimoine canadien

Canada

106-00178/

Preservation of Electronic Records: New Knowledge and Decision-making

Postprints of a Conference
Symposium 2003
Preservation of Electronic Records:
New Knowledge and Decision-making
Ottawa, Canada
September 15–18, 2003

Organized by the Canadian Conservation Institute, Library and Archives Canada, and the Canadian Heritage Information Network

Canadian Conservation Institute
Ottawa, Canada
2004

La préservation des documents électroniques : Information récente et prise de décisions

Actes du congrès
Symposium 2003
La préservation des documents électroniques :
Information récente et prise de décisions
Ottawa, Canada
Du 15 au 18 septembre 2003

Organisé par l'Institut canadien de conservation, Bibliothèque et Archives Canada et le Réseau canadien d'information sur le patrimoine

Institut canadien de conservation
Ottawa, Canada
2004

© Minister of Public Works and Government Services, Canada, 2004

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, entering in an information storage and retrieval system, or otherwise, without prior written permission of the publisher.

PRINTED IN CANADA

Available from:

Canadian Conservation Institute
1030 Innes Road
Ottawa ON K1A 0M5
Canada

Library and Archives Canada Cataloguing in Publication

Main entry under title = Vedette principale au titre

Preservation of electronic records : new knowledge and decision-making :
postprints of a conference, symposium 2003 = La préservation des documents électroniques : information récente et prise de décisions : actes du congrès,
symposium 2003

Text in English and French.
"Ottawa, Canada. September 15-18, 2003"
ISBN 0-662-68620-9
Cat. no. NM95-59/12-2005

1. Electronic records – Conservation and restoration – Congresses.
2. Digital preservation – Congresses.
3. Computer files – Conservation and restoration – Congresses.
4. Libraries materials – Conservation and restoration – Congresses.
I. Canadian Conservation Insitute.
II. Title: La préservation des documents électroniques, information récente et prise de décisions.

Z701.3C65P43 2005 025.17′4 C2004-980352-2E

Disclaimer: The opinions expressed in these articles are those of the authors, and do not necessarily reflect those of the Canadian Conservation Institute.

© Ministre, Travaux publics et Services gouvernementaux, Canada, 2004

Tous droits réservés. La reproduction d'un extrait quelconque de ce livre, par quelque procédé que ce soit, tant électronique que mécanique, ou par photocopie, microfilm, bande magnétique, disque ou autre, sans l'autorisation écrite de l'éditeur est interdite.

IMPRIMÉ AU CANADA

Vendu par :

Institut canadien de conservation
1030, chemin Innes
Ottawa ON K1A 0M5
Canada

Catalogage avant publication de Bibliothèque et Archives Canada

Main entry under title = Vedette principale au titre

Preservation of electronic records : new knowledge and decision-making :
postprints of a conference, symposium 2003 = La préservation des documents électroniques : information récente et prise de décisions : actes du congrès,
symposium 2003

Texte en anglais et en français.
«Ottawa, Canada. Du 15 au 18 septembre 2003 »
ISBN 0-662-68620-9
N° de cat. NM95-59/12-2005

1. Archives électroniques – Conservation et restauration – Congrès.
2. Numérisation – Congrès.
3. Fichiers (Informatique) – Conservation et restauration – Congrès.
4. Documentation de bibliothèque – Conservation et restauration – Congrès.
I. Institut canadien de conservation.
II. Titre: La préservation des documents électroniques, information récente et prise de décisions.

Z701.3C65P43 2005 025.17′4 C2004-980352-2F

Mise en garde : Les opinions exprimées dans les présents articles n'engagent que la responsabilité des auteurs et ne reflètent pas nécessairement les points de vue de l'Institut canadien de conservation.

Table of Contents/Table des matières

Foreword / Avant-propos i

Organizing Committee / Comité organisateur ii

Acknowledgements / Remerciements iii

Preface / Préface v

Introduction: *Choices and Decisions* 1
Tom Strang

Appraisal of Electronic Records/Évaluation des documents électroniques

Appraising Digital Materials for Preservation as Cultural Heritage 7
Terry Eastwood

Preserving Digital Documentary Heritage in Libraries: What Do We Select? 13
Ross Harvey

Selecting for Survival: Developing a Model for Selecting Audiovisual Archival Documents for Preservation Reformatting 21
Nancy Marrelli

Authenticity of Electronic Records/Authenticité des documents électroniques

Authenticity Requirements for Electronic Records 25
Luciana Duranti

Authenticity and Accessibility, Objects and Technologies: Digital Preservation Basics 31
Bruce Walton

Developing a Preservation Strategy for Electronic Records/ Élaboration d'une stratégie de préservation des documents électroniques

Is Educating Archivists Enough?: Assessing the Impact of the ACA Institute 2002, "Approaches to the Preservation of Electronic Records" 39
Amy Marshall and Wendy Duff

Preserving Digital Objects in Small- to Medium-sized Institutions 47
Bill Walker

Developing a Digital Preservation Strategy at The British Library: Application of the Handbook Preservation Management of Digital Materials 55
Deborah Woodyard and Helen Shenton

Digital Preservation: Best Practice for Museums 61
Tim Au Yeung

Net Art et conservation, quelles stratégies pour le musée? 71
Anne Laforet

Record, Play, Fast Forward: Developing Strategies for the Care of Electronic Media Art at the Art Gallery of Western Australia 77
Vanessa Griffiths

Preserving Research Data: A Time for Action 83
Charles Humphrey

Archival Science and Archival Engineering: Building a New Future for the Past 93
Kenneth Thibodeau

Digital Preservation Strategies: The Initial Outcomes of the ERPANET Case Studies 99
Seamus Ross, Monica Greenan, and Peter McKinney

Preservation Strategies for Electronic Records/Stratégie de préservation des documents électroniques

Implementation of the Victorian Electronic Records Strategy (VERS) within the Government of the State of Victoria, Australia 115
Howard Quenault

Volatile Memory 121
Ulrich Lang

Robert Creeley's Computer Files: A Collaborative Approach to link Access with Preservation 127
Michael G. Olson

Old Wine in New Skins: Migrating Legacy Data in a Museum Context 135
Richard Gerrard and David Harkness

A Strategy for Archiving Web Sites at Library and Archives Canada 143
Susan Haigh and Roselyn Lilleniit

Preserving Audiovisual Records of the University of Manitoba Archives and Special Collections 149
Jane Dalley

Preservation Strategies for Digital Collections: A Critical Review 155
Franziska S. Frey and Stephen Chapman

Computer-based Sound Recording Preservation and its Application in Small Communities 157
Norm Glowach and John Poirier

Preservation and the Principles of Access 165
Alexander Eykelhof

Archiving Electronic Media 171
Jim Wheeler

Media Knowledge/Connaissance des supports

Preservation of Electronic Records: Status of ISO Standards 181
Peter Z. Adelstein

Waiting to Happen: Lessons from Preserving Disaster-afflicted Electronic Media in an Archival Collection 189
Hannah Frost

Longevity of CD Media: Research at the Library of Congress 197
Chandru J. Shahani, Basil Manns, and Michele Youket

Restless Records: The State of Small Video Art Archives in Canada 207
Maija Martin

Posters/Affiches

Survey Results: Preservation Management of Born-digital Documents in Manuscript Repositories in the United States 217
Ruth Elizabeth Bryan

Virtual Remote Control: Preservation Risk Management for Web Resources 219
Nancy Y. McGovern, Anne R. Kenney, Richard Entlich, William Kehoe, and Peter Botticelli

Role of Lacquer and Ink in Protecting Compact Discs exposed to Gaseous Pollutants 221
Léon-Bavi Vilmont

Foreword

The preservation of electronic records is one of the latest challenges facing the conservation and heritage communities. It is a complex and multifaceted task that includes content, media, hardware, and software.

The program for *Symposium 2003 - Preservation of Electronic Records: New Knowledge and Decision-making* was developed to deal systematically and logically with the various issues. The organizing partners — the Canadian Conservation Institute (CCI), Library and Archives Canada (LAC), and the Canadian Heritage Information Network (CHIN) — all have extensive experience in the field of electronic record preservation, and each brought a unique perspective on specific aspects of the topic.

But the challenge of preserving electronic records extends well beyond the traditional heritage community. It really includes all custodians of electronic information — from corporations to government agencies to individuals. To meet this need, the symposium also included a separate half-day event for the general public.

The organizing partners were delighted to welcome more than 350 delegates to the symposium; of these, 85% were from Canada, 10% from the United States, and the rest from a variety of countries including Australia, Bermuda, the Cayman Islands, Cuba, France, Italy, Malaysia, Mexico, the Netherlands, New Zealand, Taiwan, and the United Kingdom. Everyone participated actively in the discussions, and returned to their institutions with not only a better understanding of the challenges but also with viable and practical solutions that can be implemented immediately.

Charles Costain
Acting Director General
Canadian Conservation Institute (CCI)

Ian E. Wilson
National Archivist
Library and Archives Canada (LAC)

Roch Carrier
National Librarian
Library and Archives Canada (LAC)

Jean-Marc Blais
Acting Director General
Canadian Heritage Information Network (CHIN)

Avant-propos

La préservation des documents électroniques compte parmi les plus récents défis que doivent relever les collectivités de la conservation et du patrimoine. Cette tâche complexe concerne de nombreux éléments tels que le contenu, les médias, le matériel et les logiciels.

L'objectif du programme du *Symposium 2003 – La préservation des documents électroniques : Information récente et prise de décisions* est de se pencher systématiquement et logiquement sur les divers enjeux. Les organisateurs partenaires – l'Institut canadien de conservation (ICC), Bibliothèque et Archives Canada (BAC) et le Réseau canadien d'information sur le patrimoine (RCIP) – possèdent tous une vaste expérience de la préservation des documents électroniques, et chacun d'entre eux apporte un éclairage unique à des aspects précis de la question.

Mais le défi que représente la préservation des documents électroniques dépasse de loin la collectivité du patrimoine. Il inclut en fait tous les détenteurs de renseignements électroniques – des sociétés aux organismes gouvernementaux, et jusqu'aux particuliers. En vue de satisfaire ce besoin, une activité d'une demi-journée à l'intention du grand public fut présentée dans le cadre du symposium.

Les organisateurs partenaires ont été ravis d'accueillir plus de 350 participants, dont 85 % venaient du Canada, 10 % des États-Unis et 5 % de divers pays y compris l'Australie, les Bermudes, les îles Caïmans, Cuba, la France, l'Italie, la Malaisie, le Mexique, les Pays-Bas, la Nouvelle-Zélande, Taïwan et le Royaume-Uni. Tous ont pris part activement aux discussions, et tous sont retournés dans leurs établissements avec non seulement une meilleure compréhension des défis à relever, mais également avec des solutions pratiques qu'ils seront en mesure de mettre en œuvre sur-le-champ.

Charles Costain
Directeur général par intérim
Institut canadien de conservation (ICC)

Ian E. Wilson
Archiviste national du Canada
Bibliothèque et Archives Canada (BAC)

Roch Carrier
Administrateur général / Bibliothèque nationale du Canada
Bibliothèque et Archives Canada (BAC)

Jean-Marc Blais
Directeur général par intérim
Réseau canadien d'information sur le patrimoine (RCIP)

Organizing Committee

Co-Chairs: Jane Down (CCI), Mary Murphy (LAC), Richard Green (LAC), Pat Young (CHIN)

Program: Joe Iraci (CCI) (*Chair*), Tom Strang (CCI), René Paquet (LAC), Bruce Walton (LAC), Gilles St-Laurent (LAC), Fay Hjartarson (LAC), Sheila Carey (CHIN)

Posters: Wanda McWilliams (LAC) (*Chair*), Alison Bullock (LAC), Janet Mason (CCI), Greg Hill (LAC), Bonnie Szirtes (CHIN)

Publications/Signs: Barbara Patterson (CCI) (*Chair*), Linda Leclerc (CCI), Sophie Georgiev (CCI)

Tours: Mary-Lou Simac (CCI) *(Chair)*, Brenda Muir (LAC), Mike Thompson (LAC), Denis Gagnon (LAC)

Preservation Quest: Linda Street (CCI) (*Chair*), Joe Iraci (CCI), Lise Perron-Croteau (CCI), Carol MacIvor (CCI), Shanna Ramsay (CCI), Yvette Hackett (LAC), Louisa Coates (LAC), Randall Ware (LAC), Chris Dicks (LAC), Paul Lima (CHIN), Jane Down (CCI), Greg Hill (LAC)

Local Arrangements: Greg Hill (LAC) (*Chair*), Katherine Phillips (LAC), Natalie Labonté (LAC), Judith Falardeau (LAC), Colette Landry (CCI)

Trade Show/Marketing: Shanna Ramsay (CCI) (*Chair*), Chris Dicks (LAC)

Registration/Correspondence: Christine Bradley (CCI) (*Chair*)

Finance: Colette Landry (CCI) (*Chair*)

Media Relations: Carol MacIvor (CCI) and Louisa Coates (LAC) (*Co-Chairs*), Pauline Portelance (LAC), Paul Lima (CHIN), Shanna Ramsay (CCI)

Comité organisateur

Coprésidents : Jane Down (ICC), Mary Murphy (BAC), Richard Green (BAC), Pat Young (RCIP)

Programme : Joe Iraci (ICC) (*président*), Tom Strang (ICC), René Paquet (BAC), Bruce Walton (BAC), Gilles St-Laurent (BAC), Fay Hjartarson (BAC), Sheila Carey (RCIP)

Affiches : Wanda McWilliams (BAC) (*présidente*), Alison Bullock (BAC), Janet Mason (ICC), Greg Hill (BAC), Bonnie Szirtes (RCIP)

Publications/Panneaux : Barbara Patterson (ICC) (*présidente*), Linda Leclerc (ICC), Sophie Georgiev (ICC)

Visites : Mary-Lou Simac (ICC) (*présidente*), Brenda Muir (BAC), Mike Thompson (BAC), Denis Gagnon (BAC)

À la sauvegarde de votre patrimoine : Linda Street (ICC) (*présidente*), Joe Iraci (ICC), Lise Perron-Croteau (ICC), Carol MacIvor (ICC), Shanna Ramsay (ICC), Yvette Hackett (BAC), Louisa Coates (BAC), Randall Ware (BAC), Chris Dicks (BAC), Paul Lima (RCIP), Jane Down (ICC), Greg Hill (BAC)

Organisation locale : Greg Hill (BAC) (*président*), Katherine Phillips (BAC), Natalie Labonté (BAC), Judith Falardeau (BAC), Colette Landry (ICC)

Salon commercial/Promotion : Shanna Ramsay (ICC) (*présidente*), Chris Dicks (BAC)

Inscription/Correspondance : Christine Bradley (ICC) (*présidente*)

Finances : Colette Landry (ICC) (*présidente*)

Relations avec les médias : Carol MacIvor (ICC) et Louisa Coates (BAC) (*coprésidentes*), Pauline Portelance (BAC), Paul Lima (RCIP), Shanna Ramsay (ICC)

Acknowledgements

The hosting partners would like to extend their thanks to the Organizing Committee and all volunteers, speakers, poster presenters, session chairs, exhibitors, and sponsors, including the following organizations:

Audio Conservation Studio of Library
 and Archives Canada
Audio Visual Preservation Trust
Gatineau Preservation Centre
Canada Museum of Science and Technology
Canadian Conservation Institute
Canadian Heritage Information Network
Canadian Museum of Civilization
Department of Canadian Heritage
Library and Archives Canada
The Willow Group
Translation Bureau, Public Works and
 Government Services Canada
Tunstall & Tunstall Data Recovery Services

Remerciements

Les partenaires hôtes désirent remercier le Comité organisateur et tous les bénévoles, les conférenciers, les personnes qui ont présenté des affiches, les présidents de séances, les exposants et les commanditaires, y compris les organismes suivants :

Bibliothèque et Archives Canada
L'Institut canadien de conservation
Le Bureau de la traduction, Travaux publics
 et Services gouvernementaux Canada
Le Centre de préservation de Gatineau
Le ministère du Patrimoine canadien
Le Musée canadien des civilisations
Le Musée des sciences et de la technologie du Canada
Le Réseau canadien d'information sur le patrimoine
Le studio des enregistrements sonores de la Division
 de la musique (Bibliothèque et Archives Canada)
Le Trust pour la préservation de l'AV
Tunstall & Tunstall Data Recovery Services
The Willow Group

Preface

Symposium 2003 - Preservation of Electronic Records: New Knowledge and Decision-making took place in Ottawa on September 15–18, 2003, with a program that took the form of a decision tree. Following this format, the sessions could focus on areas that related to each other in a structured manner and all the key issues relating to the preservation of electronic records could be introduced in a logical sequence from the opening to closing speakers. The topics discussed included not only storage media but also the issues that must be addressed before considering how long an electronic record will last. [For the purpose of this symposium, the term "electronic record" was used to describe a large variety of records, e.g. audio and video recordings in analog or digital form and data files such as text, spreadsheets, e-mails, etc.]

The decision tree was initially formulated by the Program Committee, but refined by Tom Strang from CCI and Bruce Walton from LAC. Addressing the key decisions/choices that an institution needs to make when considering the acquisition and preservation of electronic records led to five main session categories: Appraisal of Electronic Records; Authenticity of Electronic Records; Developing a Preservation Strategy for Electronic Records; Preservation Strategies for Electronic Records; and Media Knowledge. The call for papers clearly outlined this approach, and resulted in the submission of more than 45 abstracts. Because of the decision tree program structure, the process to select the papers was very specific. The final program consisted of 29 papers from 6 countries: Canada (14); United States (8); Australia (3); United Kingdom (2); France (1); and Germany (1). In addition to the main program, 4 posters that fit into the program structure were also presented (1 from Canada, 2 from the United States, and 1 from France).[1]

Another objective of the program was to feature a wide variety of small- to medium-sized institutions that included not only archives and libraries, but also cultural institutions such as art galleries and museums that are faced with preserving electronic records. For example, art galleries often include video art in their collections, but the needs of these electronic records are quite different than the needs

Préface

Symposium 2003 – La préservation des documents électroniques : Information récente et prise de décisions a eu lieu à Ottawa, du 15 au 18 septembre 2003, avec un programme qui a pris la forme d'un arbre de décision. Selon ce format, les séances pouvaient porter sur des sujets liés entre eux de façon structurée et tous les points importants concernant la préservation des documents électroniques pouvaient être présentés logiquement, du premier au dernier conférencier. Les sujets discutés incluaient non seulement les supports de préservation, mais aussi les questions à traiter avant de se demander quelle est la durée de vie d'un document électronique. [*Nota* : Pour les fins de ce symposium, le terme « document électronique » a été utilisé pour désigner un large éventail de documents, comme les enregistrements audio et vidéo sous forme analogique ou numérique et les fichiers de données tels que les textes, les tableurs, les courriels, etc.]

L'arbre de décision a d'abord été formulé par le comité du programme et raffiné ensuite par Tom Strang, de l'ICC, et Bruce Walton, de BAC. Les choix les plus importants que doit faire une institution et les décisions clés qu'elle doit prendre au moment de penser à l'acquisition et à la préservation des documents électroniques ont mené à l'établissement de cinq catégories principales de séances : Évaluation des documents électroniques; Authenticité des documents électroniques; Élaboration d'une stratégie de préservation des documents électroniques et Connaissance des supports. Cette approche a été soulignée dans la demande de communications, et plus de 45 résumés ont été soumis. Grâce à la structure du programme en arbre de décision, le processus de sélection a été très précis. Le programme final a consisté en 29 communications représentant six pays : le Canada (14); les États-Unis (8); l'Australie (3); le Royaume-Uni (2); la France (1) et l'Allemagne (1). En plus du programme principal, quatre affiches qui cadraient avec la structure du programme furent également présentées (une du Canada, deux des États-Unis et une de la France).[1]

Un autre objectif du programme consistait à présenter une grande variété de petites à moyennes institutions possédant non seulement des archives et des of the electronic records typically found in archives and libraries. Thus, the speakers list included several individuals from the museum and gallery communities.

---

[1] One of the posters that was presented at the symposium was the decision tree that appears on p. 2 of the "Introduction" of this book of postprints. Hence, only three abstracts are included in the "Posters" section.

[1] Une affiche présentée au symposium portait sur l'arbre de décision qui apparaît à la page 2 de l'introduction du présent ouvrage. Par conséquent, seules trois affiches font partie de la section sur les affiches.

Finally, it was important that the program include some case histories that highlighted what various institutions have actually done or are doing to preserve electronic records. While discussing preservation strategies is important, seeing strategies actually being implemented and working in the real world provides useful information and models to follow. Among the case histories presented were the preservation of audio language recordings from Aboriginal elders in the Northwest Territories in Canada, the preservation of American poet Robert Creeley's computer files, and the implementation of the Victorian Electronic Records Strategy (VERS) within the Government of the State of Victoria, Australia. The case histories were arguably the most useful aspect of a program that included something for everyone.

As a whole, the program provided delegates with a sense of the broader issues involved in collecting and preserving electronic records, as well as knowledge about the challenges that other institutions are facing and how they are dealing with them.

We are pleased to present the papers from the symposium program in this book of postprints. However, because speakers were given the opportunity to revise their papers after the symposium and the final submissions were lightly edited, the text herein may differ slightly from the original presentations.

Jane Down, CCI
Richard Green, LAC
Joe Iraci, CCI
Mary Murphy, LAC
Patricia Young, CHIN

Note: The papers in this book are published in the language in which they were presented, but all include an abstract in both English and French.

bibliothèques, mais aussi des institutions culturelles telles que des musées qui doivent conserver des documents électroniques. Les musées, par exemple, incluent souvent de l'art vidéo dans leurs collections, mais les besoins de consultation de ces documents électroniques sont bien différents de ceux qu'on retrouve habituellement dans les archives et les bibliothèques. Ainsi, la liste des conférenciers comprenait plusieurs personnes du monde muséal.

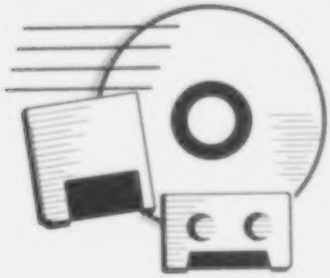
Enfin, il était important de retrouver dans le programme des histoires de cas afin d'illustrer ce que diverses institutions ont fait ou font concrètement pour préserver leurs documents électroniques. La discussion des stratégies de préservation est certes importante, mais le fait de connaître des stratégies qui sont mises en œuvre et qui fonctionnent dans une situation réelle permet d'obtenir de l'information utile et d'avoir des modèles à suivre. Parmi les cas décrits, il y a eu la préservation des enregistrements sonores des aînés autochtones des Territoires du Nord-Ouest, au Canada, la préservation des fichiers informatiques du poète américain Robert Creeley et la mise en application de la stratégie VERS pour les documents électroniques de l'État de Victoria, en Australie. On peut soutenir que les cas présentés formaient l'aspect le plus utile d'un programme répondant aux attentes d'un grand nombre de participants.

En général, les délégués sont repartis avec une vue d'ensemble des grandes questions liées à la collection et à la préservation des documents électroniques de même qu'une connaissance des problèmes auxquels se heurtent d'autres institutions ainsi que de leur manière de les résoudre.

Nous sommes fiers de vous présenter les communications du symposium dans cet ouvrage. Cependant, étant donné que les conférenciers ont eu l'occasion de modifier leurs textes après le symposium et qu'ils ont été légèrement révisés, il est possible que les textes soient quelque peu différents de ceux présentés durant le symposium.

Jane Down, ICC
Richard Green, BAC
Joe Iraci, ICC
Mary Murphy, BAC
Patricia Young, RCIP

Remarque : les communications sont publiées dans la langue utilisée lors de la présentation, mais toutes sont accompagnées d'un résumé en français et en anglais.



INTRODUCTION: CHOICES AND DECISIONS

TOM STRANG

Canadian Conservation Institute
Ottawa ON K1A 0M5
Canada
Tel.: (613) 941-4990
E-mail: tom_strang@pch.gc.ca

Abstract

This conference is aimed at helping small- to moderate-sized collections of "modern media" to survive, and is based on both practical institutional experience and current research. The program is designed around a decision tree that organizes the theoretical process, working-knowledge, and content gaps into a visual path. There are three main sections: the appraisal and pre-acquisition phase; the region entangled by the nature of physical collections; and the conceptual space where the future accessibility for information is decided. The splitting of content from carrier is a useful concept for information theorists, and is a major difference from traditional collections. Answering the question "What form could you accept for preserving your information while retaining its validity?" is the practical side of this conceptual split, and many institutions would be happy with a suitable answer to this question. There are three aphorisms that encapsulate this practical side of modern media preservation: move it or lose it; something is better than nothing; and don't throw away the original. Although there is already a spectrum of possible means to the goal of preserving authentic objects for long periods, the hazards and benefits associated with these approaches need to be determined and explained before achievable preservation plans can be proposed and chosen. Modern media collections are in jeopardy of malfunction before stable societies have traditionally expected to lose material heritage. The decision tree on which this symposium is based is offered in the spirit that it will be news to some that informed choices can be made to carry this modern media heritage forward through time.

Introduction

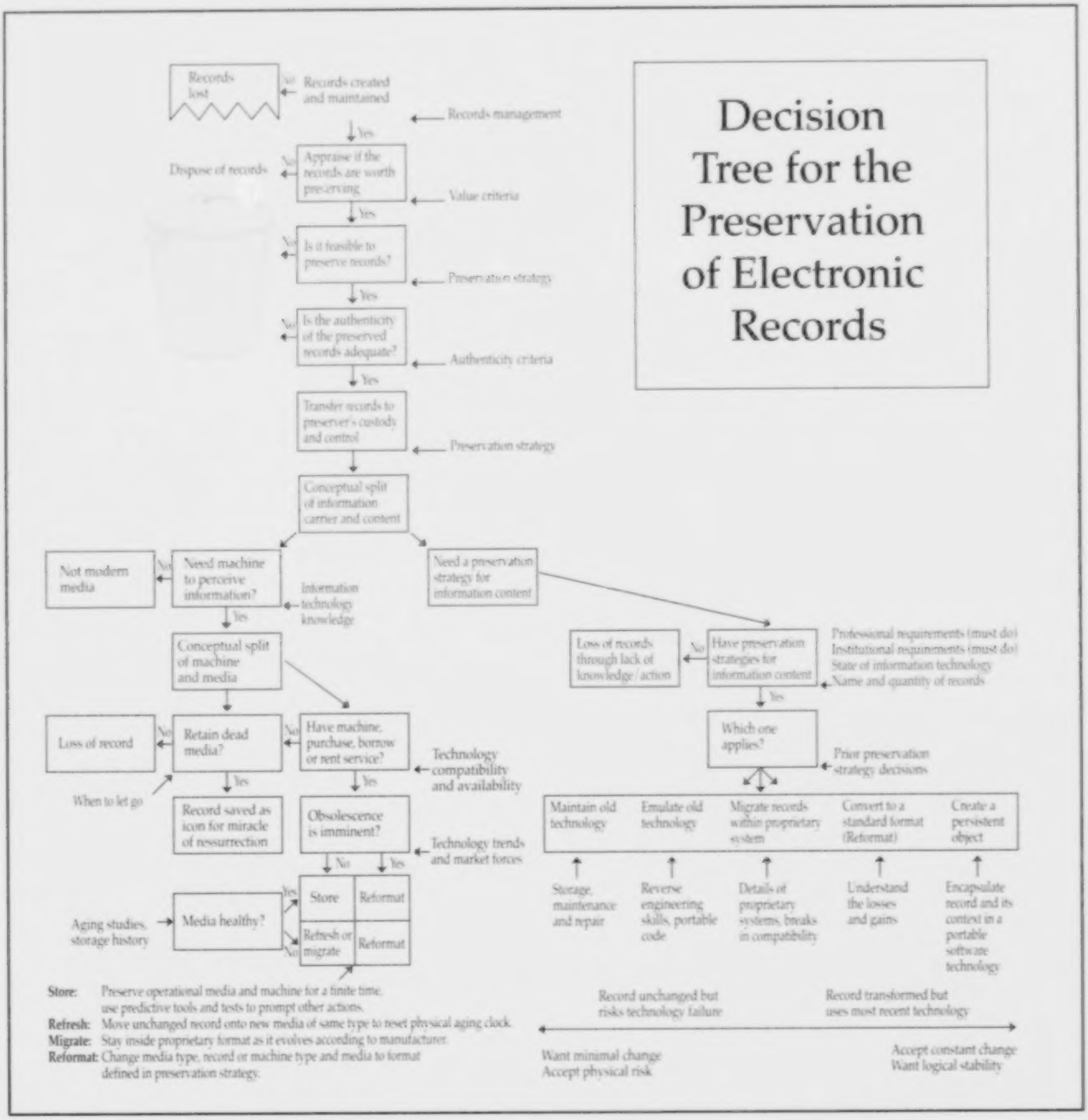
Many small- to medium-sized institutions are currently facing a problem in regards to their holdings of electronic records: how can they keep their collections accessible in the face of technology that is quickly becoming obsolete and information carriers that are rapidly deteriorating in storage conditions that seem benign. In many cases, these institutions are only beginning to come to grips with this problem as the volume of their material rapidly increases. And if the diverse history of audio and video is not enough, personal computers have contributed the final straw.

From 1999 to 2003, the Canadian Conservation Institute (CCI) ran workshops across Canada on modern information carriers in an attempt to relate the fragilities and strengths of optical disc media and magnetic tape to the archives community. The inexorable deterioration of magnetic media at ambient conditions was a message that was new to many participants. But beyond our material-centric approach, many real-world questions were asked: How do you preserve a videographer's lifetime of work on aging tape? Is transfer to MPEG and DVD an answer or a further problem? How do you help a small archive who is receiving electronic records in forms that span across several generations of obsolete computers? Is migration from audio tape to "Red Book" audio CD the best preservation strategy? Of the many questions asked, only a few could be answered.

The preservation community (a term that is meant to include archives, libraries, and museums) has naturally made progress in these areas and has some solutions for preserving modern information. This symposium is being offered by CCI and Library and Archives Canada (LAC) to provide a venue for this community to come together and discuss the topic.

The decision tree

The decision tree was developed by the symposium program committee (with special thanks to Bruce Walton, LAC) to serve as the backbone for organizing the program. With this structure in mind, the Call for Papers was then designed to elicit papers on the specific knowledge areas identified on the tree.


Decision Tree for the Preservation of Electronic Records

Records lost
No
Records created and maintained
Records management
Yes
Dispose of records
No
Appraise if the records are worth preserving
Value criteria
Yes
No
Is it feasible to preserve records?
Preservation strategy
Yes
No
Is the authenticity of the preserved records adequate?
Authenticity criteria
Yes
Transfer records to preserver's custody and control
Preservation strategy
Conceptual split of information carrier and content
Not modern media
No
Need machine to perceive information?
Information technology knowledge
Yes
Need a preservation strategy for information content
Conceptual split of machine and media
Loss of record
No
Retain dead media?
No
Have machine, purchase, borrow or rent service?
Technology compatibility and availability
When to let go
Yes
Record saved as icon for miracle of ressurrection
Yes
Obsolescence is imminent?
Technology trends and market forces
No
Yes
Aging studies, storage history
Media healthy?
Yes
No
Store
Reformat
Refresh or migrate
Reformat
Loss of records through lack of knowledge/action
No
Have preservation strategies for information content
Professional requirements (must do)
Institutional requirements (must do)
State of information technology
Name and quantity of records
Yes
Which one applies?
Prior preservation strategy decisions
Maintain old technology
Emulate old technology
Migrate records within proprietary system
Convert to a standard format (Reformat)
Create a persistent object
Storage, maintenance and repair
Reverse engineering skills, portable code
Details of proprietary systems, breaks in compatibility
Understand the losses and gains
Encapsulate record and its context in a portable software technology
Record unchanged but risks technology failure
Record transformed but uses most recent technology
Want minimal change
Accept physical risk
Accept constant change
Want logical stability

Store: Preserve operational media and machine for a finite time, use predictive tools and tests to prompt other actions.
Refresh: Move unchanged record onto new media of same type to reset physical aging clock.
Migrate: Stay inside proprietary format as it evolves according to manufacturer.
Reformat: Change media type, record or machine type and media to format defined in preservation strategy.

To assist in referencing to their content, contributed papers are grouped by their region within the tree. Beyond this immediate use, the tree was intended to serve as a simple document to convey the wider picture of the process used and attendant knowledge needed when preserving a modern information collection. Intellectually, the decision tree also owes a fair amount to the InterPARES project,[1] but in tree form it is not intended to be as comprehensive in detail. The tree has subsequently become a discussion piece for many of CCI's workshops on modern information carriers; participants have used it to elaborate on their approach to acquiring and managing electronic media records and problems that arise.

---

1. The InterPARES (International Research on Permanent Authentic Records in Electronic Systems) project was a collaborative, multidisciplinary, and international research endeavour that involved 60 researchers from 11 countries. Its aim was to develop the theoretical and methodological knowledge necessary for the long-term preservation of the authenticity of electronic records.

The tree settled into three main regions graphed as an inverted Y. The upper branch covers the process of appraisal for preservation and acquisition. Potential collections are winnowed by the desire to preserve (value), and the methods and resources to preserve. This is a two-pass filter: "If we can preserve it, we can collect it" is the major threshold; and once the decision is made to collect the material, the actual process (preservation strategy) can be enacted as envisioned. This is opposed to the alternate approach of "collect it and hope we can preserve it someday." The knowledge areas required are value criteria, preservation strategy, and authenticity criteria, and subsequent papers in this symposium reflect these areas.

The lower branches relay the conceptual split between carrier and content. The nature of electronic media, and especially the possibility of digital representation, allows one to think in terms of disembodied "information streams." However, the hard truth is that these lumps of information must reside somewhere and be maintained, searchable, and accessible; otherwise all this effort is not worthwhile. In some sense the two branches represent the way things are and the way they could be. At best, they are what you've got and what you want. At worst, they are all you've got and a constant frustration.

The leftmost branch shows basic requirements for keeping machine-readable records available for reference. While generally suggesting forestalling actions against loss, this area encapsulates in a two-by-two matrix the (minimum) possible preservation actions based on the idea of relative media and machine health. The definitions of store, refresh, migrate, and reformat are provided because these terms are jargon associated with specific choices. The knowledge areas required to make decisions about priority in processing include general recognition of media technologies, technology compatibility and availability, technology trends and market forces, aging studies, and storage history of media.

The rightmost branch follows the preservation strategy, and terminates in the spectrum of choices that have been suggested in the left branch matrix. The preservation strategy is the forward path for the collection and can entail transformation of the existing record, deliberate transcription from the acquired media, and departure from original hardware to a new corpus. This is not so difficult to envision for electronic text because we are already accustomed to viewing texts in various formats — and all are ultimately legible. But the discussion about digital forms for analog sources is a more gnawing one because we must lose some information in the sampling. So the choices are presented as a spectrum flowing from desiring minimal change but having to accept physical risk by deterioration, toward the methods that accept constant corporeal change but maintain stable and recognizable authentic records in some logical form.

In the course of CCI workshops, we have reduced the entire process into three discussion-challenging aphorisms:

- "Something is better than nothing" evokes the inherent stress of selection for preservation, and grudging acceptance of sampling analog originals as a preservation mechanism by digital means.
- "Move it or lose it" captures the urgency of some situations, and affirms the pragmatism of something being better than nothing.
- "Don't throw out the original" is our caveat. Competent transfer of "born-digital" text can safely end in disposal of the old media. Rarely are the creation of digital collections for the purpose of access confused with preservation strategies, but we want to ensure these are distinguished as different processes. Of most concern is that the choices for preservation formats are still in flux within the area of audio and especially video records. We want to ensure that the original survives to its "natural" time with extension by good storage conditions, rather than being discarded prematurely only to wish we could go back and resample.

We hope the decision tree has some lasting value as a tool for the creation of preservation strategies for modern information carriers in small- to medium-sized collections — even if only as a stepping stone to greater comprehension.

Bibliography

Michalski, S. *Guidelines for Humidity and Temperature for Canadian Archives*. Technical Bulletin, No. 23. Ottawa, ON: Canadian Conservation Institute, 2000.

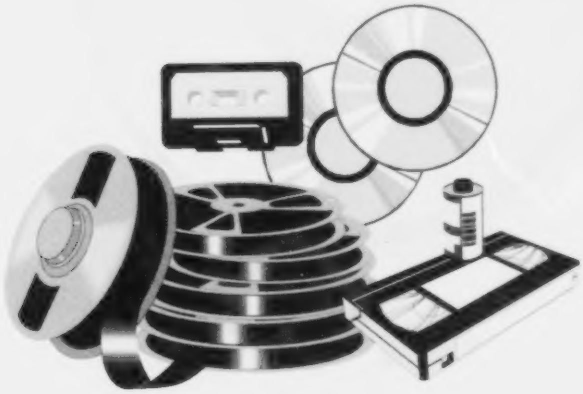
InterPARES (International Research on Permanent Authentic Records in Electronic Systems) Project. www.interpares.org/index.htm

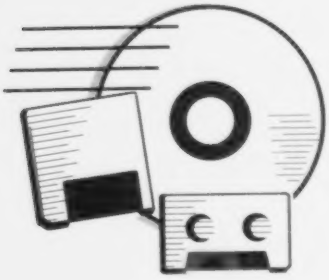
Résumé

Le Symposium a été conçu pour aider les établissements responsables de petites et moyennes collections de « médias modernes » à survivre. Il est fondé à la fois sur une expérience institutionnelle pratique et sur les recherches actuelles. La création d'un arbre de décision permet l'organisation en un schéma visuel du processus théorique, des connaissances pratiques et des lacunes sur le plan du contenu. Le programme du Symposium a été élaboré en fonction d'un diagramme de ce genre. Les trois principales sections sont les suivantes : l'étape de l'évaluation et de la pré-acquisition, celle de la nature des collections physiques, et l'espace conceptuel où l'accessibilité future de l'information est déterminée. La séparation du contenu de son support constitue un concept utile aux théoriciens de l'information, ainsi qu'une différence majeure par rapport aux collections traditionnelles. Répondre à la question : « Quelle forme êtes-vous prêt à accepter pour préserver votre information tout en conservant sa valeur? », c'est le côté pratique de cette division conceptuelle. Nombreuses sont les institutions qui aimeraient avoir une bonne réponse à donner à cette question. Trois formules rendent compte de l'aspect pratique de la préservation des médias modernes : « À transférer ou à abandonner », « Vaut mieux garder quelque chose que de se retrouver avec rien », et « Ne jetez pas l'original ». À l'heure actuelle, nous reconnaissons une gamme de moyens possibles de préserver des objets authentiques à long terme. Les risques et les avantages liés à ces méthodes doivent être précisés et expliqués, de manière à ce que des plans de préservation réalistes puissent être proposés et choisis. Les collections de médias modernes sont en danger de défaillance, et ce, bien avant le moment où les sociétés stables s'attendent habituellement à perdre leurs biens patrimoniaux. Cet arbre décisionnel est proposé dans l'intention d'apprendre à ceux qui l'ignorent qu'il est possible de faire des choix éclairés afin de préserver ce patrimoine pour les générations futures.

APPRAISAL OF ELECTRONIC RECORDS

ÉVALUATION DES DOCUMENTS ÉLECTRONIQUES





APPRAISING DIGITAL MATERIALS FOR PRESERVATION AS CULTURAL HERITAGE

TERRY EASTWOOD

School of Library, Archival and Information Studies
University of British Columbia
1956 Main Mall
Vancouver BC V6T 1Z1
Canada
Tel.: (604) 822-6326
E-mail: eastwood@interchange.ubc.ca

Abstract

This paper aims to extract lessons from archivists' experience in appraising electronic records that are likely to have wider application in the preservation of digital cultural heritage. It relies mainly on the work of the Appraisal Task Force of the InterPARES (International Research on Permanent Authentic Records in Electronic Systems) project on long-term preservation of authentic electronic records to develop a picture of the process of appraisal. It concludes that the aspects of assessment of authenticity, determination of the feasibility of preservation, and monitoring of electronic records are likely to find counterparts in attempts to appraise digital objects in other cultural heritage sectors. It also argues that the activities performed during appraisal constitute the first vital step in the process of preservation of digital materials.

Introduction

It is by now well-known that long-term preservation of materials in digital form presents both organizations and individuals with a significant challenge as they attempt to overcome the adverse effects of technological obsolescence and media fragility. The needs of contemporary societies the world over to keep all kinds of digital objects makes this a problem in virtually every aspect of human endeavour, but it is most pronounced for institutions consigned responsibility for preserving objects of cultural heritage — which are normally kept for as long as they can be preserved. The difficulties of long-term preservation of digital objects grow in complexity with the complexity of digital technology. These difficulties, even for relatively simple digital objects, place a premium on the capability to identify those materials of enduring value, and demand the development of a process of evaluation or appraisal adapted to the needs of the digital environment. At its essence, appraisal involves making a judgment or estimation of the worthiness of continued preservation. This paper reports on the experience and research in the archival community in coming to grips with this problem of appraising digital objects. It argues that this kind of appraisal is in some ways significantly different from that for traditional documentary objects, that the process is vital to effective long-term preservation, and that there is more to the exercise than meets the eye. The aim is to extract from archival experience lessons likely to be worth pondering in other cultural heritage sectors.

The appraiser as active agent of preservation

All materials preserved as objects of cultural heritage were originally produced in the course of some human activity and derive their meaning from the context of their creation. Archivists appraise, acquire, preserve, and make accessible documents made and received in the course of the affairs of individuals and organizations, and regularly communicate knowledge about them and their context as part of the exercise of ensuring their continuing accessibility. By the 1980s, archivists recognized that digital documents (which they usually call electronic records) were being produced with little regard to the requirements for their long-term preservation. By the end of that decade, it had become evident that it was difficult if not impossible to preserve digital or electronic documents produced in so-called legacy systems operating just a few years before, because too little attention had been paid to identifying important facts of their creation, relationships, and context, including their technological context. This was the era that John McDonald has called the wild frontier, when people employed computer hardware and software to make documents, store them, and maintain them with almost no thought for their use outside the immediate environment in which they were created.[1] Archivists concluded that record-keeping requirements, including provisions for determining the disposition of records, had to be part of the design of systems software. In the last few years such software has appeared, but it has not invaded all sectors of human endeavour nor has it been adequate in all situations for all the kinds of digital objects, data, and information being created today. It is likely that archivists' experience will parallel that in other spheres where digital objects are being created and need long-term preservation. In particular, two things are evident. First, digital objects have to contain the means to identify them and their context of creation so that they can be managed beyond the active system in which they were generated, and they have to be maintained so as to avoid corruption or loss of their essential characteristics. Second, the determination of their disposition (i.e. their appraisal) cannot, without significant risk, be delayed to some time in the future when the technology of creation is no longer extant.

This archival experience suggests that anyone acquiring digital objects as cultural heritage will have to seek materials actively in the here-and-now and be prepared to educate creators of these materials about the needs of long-term preservation. Seeking materials in the here-and-now is, given the proliferation of digital objects, an almost overwhelming task. Intervening with advice to promote the needs of long-term preservation is, practically speaking, a task it is all too easy to regard as being quixotic. Still, it is likely that the choice will be between actively appraising materials close to the time of their creation, or finding it is too late to consider long-term preservation or effectively bring it about.[2]

The process of archival selection

In recent years, archivists have done a lot of work to understand and characterize the various activities performed during the archival appraisal process. Relying on this body of work and on the experience of archival institutions, the Appraisal Task Force of the InterPARES project[3] on long-term preservation of authentic electronic records produced a model of the archival selection function.

The model takes the perspective of an entity responsible for long-term preservation of electronic records (hereafter called the preserver). Selecting electronic records involves appraising them and carrying out their disposition, i.e. effecting the transfer of custody of records selected for preservation from the entity creating them (hereafter called the creator) to the preserver. In any given case, the preserver needs to establish, implement, and maintain a framework of policies and procedures guiding the selection function, such as an acquisition policy or procedures governing actual physical transfer of records from the creator or, if not the creator, then from the entity having custody of them to the preserver. Preserving records responds to the creator's needs or broader societal needs to have continuing access to the records, or to both needs. However, it is particularly necessary to preserve electronic records in such a way that their identity is known and their integrity is not impaired, in short that they are authentic and are what they purport to be. The framework of policies and procedures, the creator's and societal needs, and the dictates of archival science all act as constraints on the process, as do the terms of laws or regulations that apply to records.

Archival appraisal of electronic records

Appraisal of electronic records comprises four distinct activities:

- compiling and analysing information about the records and their contexts
- assessing their capacity to serve the needs of their creator and society
- determining the feasibility of preserving them
- making the appraisal decision on the basis of the foregoing

Although evaluating digital objects other than those of archival nature for the purposes of long-term preservation may very well take on a somewhat different cast from that for records, it is worth looking at these four activities because they are likely to be very generally applicable to appraisal of all digital objects.

Compiling information

The first activity involves compiling information about the digital object or objects. Archivists normally appraise records in the natural aggregations to which they belong. These aggregations are natural in the sense that any given activity will produce records as part and parcel of conducting it. One can construe information relevant to the process of evaluation from the objects themselves about their form, content, and so on, or one can compile information from sources external to the records about the various contexts of the objects relevant to their evaluation.

The InterPARES project identified five distinct contexts of records:

1. The *juridical-administrative* context refers to the legal and administrative system in which the creating body exists. It is the broad societal context of the country or territory where the objects were produced at the time they were produced.

2. The *provenancial context* refers to the creating body, and its mandate, structure, and functions. For records produced by individuals, provenancial context would be the sphere(s) of affairs and the various activities in which the person who produced the records was involved.

3. The *procedural context* refers to the business procedures in the course of which the records were created. Although business procedures suggest a bureaucratic environment, the concept can be extended — for instance to encompass procedures in scientific activity or the routines followed in the mode of creation or production in artistic activity.

4. The *documentary context* refers to the structure and interrelationships of the whole body of records produced by an entity. It is an indispensable condition of archival appraisal that decisions take into account the whole body of records, the meaning of individual records being dependent on their various relationships with other records in the whole body of records under consideration.

5. The *technological context* refers to the technical components of the electronic systems in which the records were created — the hardware, software, data, and systems with which or within which the records were created.

Assessing value

Throughout the process of compiling contextual information, archivists also accumulate inferences about the likelihood the records under consideration will have continuing value, and about the degree to which they are likely to have remained uncorrupted or unchanged over time (i.e the degree to which the records can be presumed to be authentic). Continuing value simply refers to the capacity of records to serve the continuing interests or needs of their creator and society. For the broader purposes of this essay, continuing value might be considered the capacity of digital objects to serve as an expression of cultural heritage, or some such general notion. For electronic records (and perhaps for other digital objects) assessing value comprises assessing continuing value and assessing authenticity. Assessment of continuing value is, of course, the most critical element, but that assessment is obviously affected by an assessment of the grounds for presuming the records to be authentic. In large measure, archivists gather and evaluate evidence of what has happened to records during the course of their existence in order to assess authenticity. As the report of the Appraisal Task Force puts it:

> In cases where the chain of custody and preservation has been broken or where migration has resulted in missing records, missing parts of records, or inadequate documentation of changes, there may be good reason to suspect the value of the records. If the appraiser has good reason to suspect that the records no longer reflect what they were at the time of their creation and primary use, he or she may decide not to preserve them.[4]

Such evidence is not always easy to come by long after the fact of creation, which is one reason to intervene soon after digital objects are created. One might reasonably assume that assessment of the authenticity of other than archival digital objects will also affect judgments of continuing value. The proposition would go something like this: the greater the extent to which you can presume something to be the real thing, to use a common phrase, the more likely one is to accord it continuing value. Given the great facility that exists to alter things in the digital environment, this two-faceted assessment to determine value is very likely to become a regular feature of the process of evaluation of any kind of digital material for the purposes of long-term preservation.

Determining the feasibility of preservation

Another aspect of appraisal in the digital environment is determining the feasibility of preservation. Archivists have, of course, always had to consider the feasibility of preserving records, but for electronic records this aspect of the process becomes a more critical element in the process than it ever was with traditional materials (where it had rarely been the case that it was simply not possible to preserve). This very outcome is more likely to occur with digital materials.

To explain what is involved in determining the feasibility of preservation, suppose that you have compiled all the information you need, analysed it carefully to make an informed judgment about continuing value, and assured yourself that there is good reason to believe that what you have assessed is indeed authentic. You still have to ask yourself "Can I preserve these digital objects in such a way that the essential elements conferring their identity and ensuring their integrity can be preserved?" This question is important because, in the digital environment, records are not actually preserved but rather digital objects and the capability to reproduce them. This is not a problem so long as the digital objects exist in their native environment, but short of perpetuating that environment or re-creating it in the future, the only other strategies are to sustain the capability of making copies. This involves determining that the digital components conferring the identity and ensuring the integrity of the things you want to preserve *can indeed be preserved given current and future anticipated preservation capabilities*.

The InterPARES Appraisal Task Force broke this activity of determining feasibility of preservation into three phases:

- determining both the record elements containing informational content and those elements needing to be preserved according to the requirements for authenticity
- identifying where these crucial record elements are manifested in digital components of the electronic record that must be preserved
- reconciling these preservation requirements with the preservation capabilities of the entity that is responsible for the continuing preservation of the body of records being preserved

The InterPARES project developed a set of requirements for authenticity for archival documents or records in the digital environment. Although requirements to ensure authenticity will vary for other kinds of digital documents and objects, there seems to be little doubt that appraisers of such documents and objects will have to perform a parallel activity of determining feasibility. As the final report of the Appraisal Task Force puts it, "this feasibility determination gathers and records technical information that is necessary to accomplish preservation of the individual elements conveying both the intellectual content and the authenticity of electronic records being appraised." As well, this part of the process gathers information about "the projected cost of preservation and an indication of whether or not the preserver has [or can foresee having] the capability to preserve the records in question."[5]

At this juncture, something needs to be said about the cost of preservation. Except to mention that it is a factor in determining feasibility, the Appraisal Task Force did not dwell on cost. My own view is that the cost of preservation of digital objects of all kinds is both difficult to determine and will be rather more determinative of the outcome of appraisal than it was for traditional materials. Again, much as in the case of continuing value, there is little experience to go on. However, it is reasonable to project that some cost will be incurred regularly to cope with technological obsolescence, as well as the normal storage and maintenance costs, including copying to new media. Because these costs are difficult to predict and will probably have an escalating quality to them the further one removes oneself from the native technological context, there is bound to be extra pressure on appraisers to select only those digital objects with continuing value that can be strongly justified. When the lack of precedent for establishing continuing value is combined with the need to justify it as benefit in a cost–benefit analysis as part of determining feasibility of preservation, it is likely that calls for reappraisal will not be long in coming. In my view, much as I oppose reappraisal as a rule, it is probably necessary to accept that we are in an era of experimentation, and will have to test the efficacy of decisions we make against the cost of continuing preservation, at least until there is some solid experience on which to rely.

There is one other element of cost to consider. It seems pretty clear that developing and maintaining what almost by definition must be a cutting-edge capability to preserve digital objects is very expensive. Few archival institutions in the world have yet developed that capability, in part, no doubt, because any serious effort is costly both to establish and to keep up. It is easy enough to say that every archives, library, and, no doubt, many museums and other institutions should get into the game. Still, I think that the significant expense holds many institutions back as much as, if not more than, other factors. I would think that a national strategy of preservation of digital cultural heritage would be the best way to address these cost concerns. In the archival field, the path we have been following up to this point is not resulting in widespread preservation of electronic records. Given that information technology is widely used to create and maintain records in both formal organizational circumstances and in personal affairs, to continue on the same path risks serious loss of our ability to sustain a vibrant archival documentary heritage.

Making the appraisal decision
After carefully considering the foregoing, a final decision can be made on if, and how, electronic records should be preserved.

Monitoring appraised records

At the beginning of this paper, it was remarked that one of the earliest conclusions archivists reached about appraising electronic records was that, to be effective, it had to be done close to the time of creation or at least while the records were still active. It can be further supposed that if the records were appraised long before they were actually taken into the custody of the preserver, it will be necessary to monitor what is happening to them from time to time. The purpose of monitoring appraised records is to determine whether the terms of an appraisal of records earmarked for continued preservation are still valid. Although many changes affecting records will require little or no change to an initial appraisal, in some cases minor revisions (e.g. to the terms and conditions of transfer from the creator to the preserver) will be needed. "However, in cases where the business processes and related computer systems are significantly revamped or rebuilt, it will obviously be necessary to consider initiating a disposition under the terms of the original appraisal *and* — for the two will likely go together — redoing the appraisal to take into account the radically altered situation."[6] The problem this kind of monitoring addresses is most familiar to archival preservation tied to the ongoing production of records of a parent organization. Of course archives are not unfamiliar with acquiring records from individuals and organizations in the community. In such circumstances, appraisal and its monitoring aspect will undoubtedly be more complicated, but, in light of how difficult it is to arrive on the scene years after the fact of creation of digital material and conduct appraisal, some regimen of regular contact with prospective donors appears to be necessary in this realm as well.

Carrying out disposition of electronic records

The activity of effecting disposition of electronic records according to the appraisal decision breaks down into three activities:

- preparing electronic records for disposition
- preparing electronic records for transfer to the responsibility of the preserver
- transmitting electronic records to the preserver

As archivists see the process, appraisals set out the terms and conditions of transfer of records from the creator to the preserver, and indicate who is to do what and when they will do it.

The first step comprises copying and, if necessary, formatting those electronic records selected for preservation so as to prepare them physically for transfer. The next step is to associate records selected for preservation with the necessary information for their continuing preservation, such as the terms and conditions of transfer, identification of the digital components to be preserved, and associated archival and technical documentation needed for their treatment. This information is usually compiled and recorded during the various stages of appraisal and monitoring. "The task at this stage is to extract the information necessary for continuing preservation of the records from the mass of appraisal documentation, and associate it with the records."[7] The third step is to transmit the records selected for preservation along with accompanying information to the preserver.

Careful work in this phase of the process is absolutely vital. Archivists who are familiar with efforts to effect disposition of electronic records can recite a litany of horror stories. The wrong records have arrived, or records have arrived in the wrong format or with no accompanying documentation. Imagine turning up a disk in someone's desk drawer years hence simply because some well meaning soul asked for a copy of something interesting.

Conclusion

Preservation of electronic records and, arguably, other digital objects depends on effective appraisal (in the ways outlined in this paper) to a greater extent than does preservation of traditional materials. Although archival selection undoubtedly differs in some of its particulars from selection in other heritage preservation sectors, anyone appraising and acquiring digital objects will have to go through similar steps or phases in order to ensure their effective preservation. In this sense, appraisal is a necessary first step in the preservation process, or so it would seem to judge by the experience of archivists.

Endnotes

1. McDonald, J. "Managing Records in the Modern Office: Taming the Wild Frontier." *Archivaria* 39 (Spring 1995), pp. 70–79.

2. The conclusions reached in this section are distilled from a survey of the literature on archival appraisal done by the InterPARES project [see Eastwood, T. *Appraisal of Electronic Records: A Review of the Literature in English* (a report for the Appraisal Task Force with contributions from Shadrack Katuu, Jacqueline Killawee, and Jeff Whyte), May 2000]. www.interpares.org/documents/interpares_ERAppraisalLiteratureReview.pdf

3. The InterPARES (International Research on Permanent Authentic Records in Electronic Systems) project was a collaborative, multidisciplinary, and international research endeavour that involved 60 researchers from 11 countries. Its aim was to develop the theoretical and methodological knowledge necessary for the long-term preservation of the authenticity of electronic records.

4. *The Long-term Preservation of Authentic Electronic Records: The Findings of the InterPARES Project. Part Two — Choosing to Preserve: The Selection of Electronic Records*, p. 10. www.interpares.org/book/interpares_book_e_part2.pdf

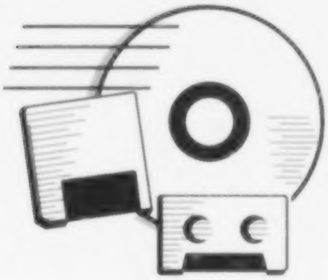
5. *Ibid.*, p. 11.

6. *Ibid.*, p. 14.

7. *Ibid.*, p. 17.

Résumé

Nous tirerons les leçons de l'expérience des archivistes en matière d'évaluation des documents électroniques, susceptibles d'avoir une application plus vaste dans la préservation du patrimoine culturel numérique. Nous nous sommes surtout inspirés du travail du Groupe d'évaluation du projet InterPARES (International Research on Permanent Authentic Records in Electronic Systems) concernant la conservation à long terme de documents électroniques authentiques pour décrire la procédure d'évaluation. Celui-ci conclut que les aspects de l'évaluation relatifs à l'intégrité, à la détermination de la faisabilité de la préservation et au suivi des documents électroniques se retrouvent probablement dans les questions que soulève l'évaluation de produits numériques dans d'autres secteurs du patrimoine culturel. Il fait également valoir que les activités d'évaluation sont une première étape cruciale de la procédure de préservation des produits numériques.



PRESERVING DIGITAL DOCUMENTARY HERITAGE IN LIBRARIES: WHAT DO WE SELECT?

ROSS HARVEY

School of Information Studies
Charles Sturt University
Locked Bag 675
Wagga Wagga NSW 2678
Australia
Tel.: +61-2-6933 2369
E-mail: rossharvey@csu.edu.au

Abstract

Libraries have long acknowledged their responsibility for preserving documents for future use. However, to ensure that documents in digital form are also preserved, the criteria that libraries have traditionally applied to selecting documents for preservation (taken from their primarily paper-based collections) need to be revisited. This paper examines traditional library selection criteria and establishes five key criteria. Emerging frameworks for selecting digital documents are also discussed, incorporating some new selection criteria and revised weightings for established criteria. The potential for the application of appraisal concepts (a practice that is fundamental to archival practice) to library preservation selection practice is suggested and measured against some of the types of digital objects managed by libraries.

Introduction: What are the issues?

Libraries need more effective guidelines for determining which documents should be retained for the future. This applies particularly to documentary heritage material in digital form, which is more vulnerable than material in traditional formats such as paper. Major resourcing issues are attached to the decisions made about which documents will be retained.

This paper is prompted by two imperatives of digital documentary heritage. [Note: In this paper the term "documentary heritage" is used to refer to any item that is movable, made up of signs/codes, sounds, and/or images, preservable (the carriers are non-living), reproducible, migratable, and the product of a deliberate documenting process (definition taken from Edmondson 2002, p. 8); and "digital object" is used to describe any object that can be represented by a computer (definition taken from Cedars Project 1999).]

The first imperative is that digital objects force us to preserve them on their own terms (Cloonan 2001, p. 237). Traditional preservation strategies, techniques, and paradigms developed in a pre-digital era do not always apply in the digital environment. For example, writing digital information to high-quality CD-R might ensure that the information will maintain its integrity into the future, but it takes no account of whether the standards used to write that information to CD-R will ensure that the information remains accessible as equipment changes. We need, therefore, to develop appropriate preservation strategies, techniques, and paradigms for the digital environment.

The second imperative is the need for conscious preservation decisions to be made about digital objects at an early stage in their existence. Burrows (2000, p. 152) contrasts the pre-digital environment with today:

> In the pre-digital era, selection necessarily preceded preservation. Once a book or journal had been acquired, a decision could be made about its long-term retention. The mere act of placing it in a library was often seen as a sufficient method of preservation in itself. In the digital era things are very different. 'Digital records don't just survive by accident', observes Margaret Hedstrom. As a result, a lack of preservation is tantamount to de-selection.

This paper attempts to contribute to thinking about selection. It first summarizes the selection criteria traditionally used by libraries when they select for preservation; it next identifies the main characteristics of the frameworks currently being developed and debated for selecting digital objects for preservation; and finally it looks at whether or not aspects of the archives community's appraisal theory and practice can assist librarians in developing more effective policies. No distinction is made between "born-digital" objects, and objects digitized as part of a digital library project or similar initiative.

Libraries and selection for preservation

Libraries have traditionally assumed responsibility for preserving some categories of documentary heritage for the future by developing and maintaining collections of physical artifacts. Their actions are guided by selection policies that are aimed at meeting the needs of their user communities. The responsibility to preserve collections of physical artifacts is much more strongly emphasized in the mission of some types of libraries (e.g. national libraries and research libraries) than of others. Although selection has always been part of a librarian's activities, many issues remain unresolved, e.g the basis on which selection is done, and the implications for current and future users — whoever they may be. The selection process requires strategies for finding out what information of significance is being created, and deciding how much of this should be kept and by whom.

The overriding imperative guiding selection decisions in many libraries is current user needs (I am excluding national libraries — who use legal deposit mechanisms to preserve comprehensive collections of national documentary resources). A secondary principle that may influence these decisions is a commitment to developing or strengthening a particular subject area, even though user needs are less immediate. But guidelines for selecting material to meet current user needs do not automatically transfer across to selecting material for retention in the future. On what basis do libraries select for preservation?

The library community's statements of principles about selection for preservation are based largely on selection of artifacts. A summary of these principles (taken from CLIR 2001, pp. 9 and 11) is provided below:

> The library preservation community has agreed on certain cardinal features of physical objects that warrant preservation in their original formats. These features are:
>
> - age
> - evidential value
> - aesthetic value
> - scarcity
> - associational value
> - market value
> - exhibition value
>
> Objective criteria or established practice determine many of these features, and the criteria vary little among libraries. They are, in short, best practice....

The key selection criteria in the above statement are:
- evidential value (evidenced by originality or fixity)
- aesthetic value (evidenced by technique and artistry of physical form)
- market value (evidenced by age, scarcity, and significance)
- associational value (evidenced by originality or fixity)
- exhibition value (strongly associated with evidential and aesthetic value)

These key selection criteria correlate well with criteria noted elsewhere, as indicated in Table 1.

Other examples provide additional criteria that might usefully be considered.

The selection criteria for the *Memory of the World* register introduce the following three (Edmondson 2002, pp. 22–23):
- significance (whether or not the item is unique and irreplaceable — something whose disappearance or deterioration would constitute a harmful impoverishment of the heritage of humanity)
- threat (whether or not its survival in danger)
- management plan (whether or not there is a plan that reflects the significance of the documentary heritage)

Harris's summary of selection for preservation in the library context suggests these three criteria (Harris 2000, pp. 206–224):
- past use (whether or not the item has been heavily used — current use being posited as an indicator, albeit imperfect, of future use)
- copyright constraints (whether or not there are any constraints caused by copyright laws)
- fragility (whether or not the item is in danger of being lost)

We can, then, characterize the approach traditionally taken by libraries to select material for preservation as being largely based on the preservation of *items* (or artifacts) in their original formats, with selection of items based on five key criteria. Some additional criteria that may be applied include physical condition, resources available, use, and social significance (an ill-defined but nonetheless well-represented category to determine if the item is "held in community esteem"). However, it must be noted that this last group of criteria is not consistently recognized by all commentators as valid.

Table 1. Key selection criteria mapped against three examples

Key selection criteria	Example 1. Selecting material for special collections (Pacey 1991, p. 189)	Example 2. Determining if material is of *national* significance (Lyall 1993, pp. 72–74)	Example 3. Selecting artifacts in museums and art galleries (Russell and Winkworth 2001, p. 11)
Evidential value	• manufacturing or evidential value • age or date of publication	• association with significant people, places, buildings, objects, events, or developments • rarity in terms of informational content • rarity in terms of format and/or recording technique • quality in terms of format and/or recording technique • quality in terms of creative and/or intellectual content • significant example of format and/or recording technique • legal importance	• scientific or research significance • provenance • representativeness • condition, intactness, and integrity
Aesthetic value	• aesthetic or artistic value	• quality in terms of format and/or recording technique • quality in terms of creative and/or intellectual content • significant example of format and/or recording technique	• aesthetic value
Market value	• rarity or uniqueness, e.g. only copy in a country • historical printing significance, e.g. first editions • cost, replacement or assessed value	• rarity in terms of informational content • rarity in terms of format and/or recording technique	• rarity
Associational value	• historical association	• association with significant people, places, buildings, objects, events, or developments	• historic significance
Exhibition value			• interpretive potential
Additional criteria	• fragility or condition	• representation and availability in public collections • physical condition if more than one copy is identified as significant	• social or spiritual significance (whether or not the artifact is "held in community esteem")

Preserving digital objects: Frameworks for selection

The traditional library preservation paradigms, developed largely to apply to paper-based objects, do not transfer well to the digital environment. Most crucially, the time frame within which decisions about the future of digital objects must be made is shortened because of the pressures of issues such as media instability and technological obsolescence. Much traditional preservation decision-making is driven by the physical condition of artifacts (Gertz 2000, p. 98) but this factor assumes little or no importance when applied to digital objects.

"Our traditional response" to deciding what to preserve has been "to keep everything that we can and wait for the significant information to rise to the surface as time, people and events dictate its importance" (Howell 2000, p. 129). This approach has been applied in the digital environment, most notably by the Internet Archive (www.archive.org). Whether this will be sustainable as the amount of digital information increases over time is debatable. Such an approach avoids the real issue of determining significance.

Much of the literature published about selection of digital information is about selecting material to digitize. This is not the same as selecting digital objects to preserve: the primary motivation of selection for digitization is to improve access to material which can be digitized, whereas selection for preservation relates principally to long-term retention of access to information already in digital form. Digitizing programs are access-driven, with preservation as a byproduct (Gertz 1999).

Despite this emphasis on access, it is useful to look at the criteria being identified for selecting materials for digitization. Many of the traditional selection criteria apply, but with different emphasis — greater prominence being given to user demand, intellectual property rights, and technical ability to preserve (Gertz 2000, pp. 98–99).

The most frequently cited criteria can be summed up as (Gertz 2000, p. 104):
- Does the item or collection have sufficient value to and demand from a current audience to justify digitization?
- Do we have the legal right to create a digital version?
- Do we have the legal right to disseminate it?
- Can the materials be digitized successfully?
- Do we have the infrastructure to carry out a digital project?
- Does or can digitization add something beyond simply creating a copy?
- Is the cost appropriate?

One constructive statement about selection for preservation of digital objects is the Cedars Project Team report (Cedars Project Team 2002). It presents the concept of a digital object's *significant properties*, defined as "the level of content and functionality retained." These significant properties are not empirical and judgments need to be made by organizations about how they relate to their user communities and their preservation responsibilities. The report proposes the example of a text in pdf. The judgment is made that the text, not the format, is significant, so information about the pdf format does not need to be stored (Cedars Project Team 2002, p. 15). The report places a high importance on intellectual property rights, which need to be negotiated before any other preservation actions occur. It suggests that selection decisions for digital preservation must currently "be pragmatic" and need to be based on the "estimated value of the material, the cost of storage and support mechanisms, and the production of metadata to support the material" (Cedars Project Team 2002, p. 53). The following specific primary criteria are proposed (Cedars Project Team 2002, section 4.5.1):
- current high use
- material (typically commercially published scholarly works) that would likely be preserved if it were published in traditional printed format
- tied to the long-term or cultural interests of the organization

Additional criteria that are proposed in a typology of digital resources (Cedars Project Team 2002, pp. 109–110) include:
- legal and intellectual property issues (legal status, i.e. intellectual property rights may need to be negotiated for preservation purposes)
- format issues (some formats are harder to preserve than others)
- technical issues (technical capacity to preserve may be lacking in a library; some technical environments may be easier to preserve than others)

A second informative statement is the *Decision Tree for Selection of Digital Materials for Long-term Retention*, in the Digital Preservation Coalition's handbook on the preservation of digital materials (Jones and Beagrie 2001, section 4). The first group of questions in this decision tree relates to selection *policy*:
- is there an institutional selection policy?
- does the material fit into it?
- is the material of significant long-term value?

The second group of questions is about *legal and intellectual property issues*:
- have acceptable rights been negotiated?
- can they be?

The third group is *technical* questions:
- can the file format be handled, now and in the future?
- can the material be transferred to a more manageable format?

And the fourth group concerns the existence of *documentation and metadata*:
- has sufficient been supplied?

To accommodate digital objects, changes are required to the selection procedures developed for traditional (non-digital) artifacts. The current approach taken by libraries is still based on the concepts developed and applied to traditional artifacts, but with an increasing realization that some factors assume greater importance, and that new factors must be incorporated. Digital selection frameworks, while still according high priority to criteria for long-term value, also place strong emphasis on other criteria: legal and intellectual property rights governing a resource, whether we have the technical ability to preserve, the costs involved in preserving, and the presence of appropriate documentation and metadata. Table 2 summarizes these changes.

Archival practice and appraisal theory

Archivists have developed a significant body of theory and practice to address the question of selection for preservation. It may be of some benefit to identify whether or not any of this theory and practice can add to the way in which librarians select digital objects for preservation. What does archival thinking add? Archivists have well-developed guidelines and practices for identifying which records should be retained and for how long. The application of these guidelines and practices is called *appraisal*, a term that lies at the core of archival practice.

Traditional definitions indicate that appraisal is the process of evaluating records to determine which are to be retained as archives, which are to be kept for specified periods, and which are to be destroyed (Ellis 1993, p. 461).

Gilliland-Swetland has demonstrated clearly that archival theory and practice can assist the development of effective practice in the digital environment. She notes (Gilliland-Swetland 2000, p. 2):

Table 2. Summary of key selection criteria

	Traditional library selection criteria	**Additional library-based criteria**	**Criteria applied to selection for digitizing**	**Cedars, Digital Preservation Coalition *DecisionTree***
Value	• evidential • aesthetic • market • associational • exhibition	• evidential • aesthetic • market • associational • exhibition	• sufficient value to current audience • added value	• significant long-term value
Physical condition		• threat to object • fragility		
Resources available		• management plan	• available infrastructure • cost appropriate	
Use		• heavy use	• current demand	• current high use
Social significance		• held in community esteem		• tied to long-term interests of organization
Legal rights		• copyright	• rights to digitize • rights to disseminate	• legal status • intellectual property rights
Format issues			• capable of being digitized successfully	• type of material (capable of being digitized successfully)
Technical issues				• technical ability to preserve available • file format can be handled
Policies				• selection policy
Documentation				• sufficient is available

> The archival perspective brings an evidence-based approach to the management of recorded knowledge. It is fundamentally concerned with the organizational and personal processes and contexts through which records and knowledge are created as well as the ways in which records individually and collectively reflect those processes. This perspective distinguishes the archival community from other communities of information professionals that manage decontextualised information.

The concept of *context* is fundamental to archival practice. Cox suggests that "from an archival perspective, context is crucial to understanding information or evidence (in any form)" (Cox 2002, p. 53) and that "records' essential elements [are] their warrant, structure, content, and context" (Cox 2003, p. 301). The emphasis placed on *context* is the key difference between library practice and archival practice in selection.

The implications of this for selecting digital objects for preservation in the library context are perhaps that *context* may be much more important than library practice has assumed, especially for digital objects. Digital objects are, by virtue of their characteristics, taking on more of the attributes of records: for example, preservation decisions need to be made about them at or near the time they are created. Librarians have already decided that the existence of documentation and metadata about digital objects, which are both aspects of context, are significant criteria to be applied to selecting digital objects for preservation. Might there be further insights to be gained from a closer scrutiny of appraisal concepts?

The brief statement about appraisal by the Archives Association of Ontario (Archives Association of Ontario 2001) assists our thinking because it provides criteria in the form of six questions:

- What is the administrative, evidential, or informational value of the records to the organization?
- Do the records meet the terms of the mandate and acquisition policy?
- Are the records primary or unique?
- Is the information in the records duplicated in another set of records?
- Can the records be properly preserved?
- Can the records be made available?

The fifth and sixth questions specifically raise the issue of practicality: are the resources (technical, financial, and human) available to preserve the records? Are access restrictions so rigid that the records cannot be made available to users in a realistic and timely way? These two criteria place a greater emphasis on the pragmatic issues than we typically find in library-derived statements. The first criterion points out the core of archival appraisal by noting the administrative, evidential, and informational value of the records to the organization. This lies at the heart of appraisal — the context aspect again, as indicated in the words "value of the records to the organization." Richard Cox (2001, p. 6) explains this further: "A record is a specific entity. It is transaction oriented. It is evidence of activity (transaction), and that evidence can only be preserved if we maintain content, structure, and context." How might we interpret this in relation to the kind of information usually handled by libraries, which are not records of transactions? Cox (2001, p. 6) provides a key: "Structure is the record form. Context is the linkage of one record to other records. Content is the information, but content without structure and context cannot be information that is reliable."

These concepts help us to reconfigure appraisal for a library preservation context. If we apply these characteristics of appraisal, developed for records, to a library context, what happens? In the library context, what are the digital object equivalents of content, structure, and context? To answer this we need to identify more clearly what kinds of digital objects libraries deal with, so that we can identify how they differ from records in the archives sense.

Towards a new decision-making framework

Examining examples of digital objects that are likely to be managed in libraries may help us to see whether the "content, structure, and context" notion could assist us in developing more precise preservation selection criteria applicable to digital objects.

The first example is of videotapes maintained by a government-owned broadcasting company's library. Information content is significant, as this broadcaster produces much material of national heritage significance. The videotapes exist in a number of formats, so technological obsolescence of recording and playback equipment is a major issue. Selection criteria need to consider the "obsolescence rating" for each format, such as *lower risk* for formats such as VHS, *vulnerable* for Digital Betacam (Sony), *threatened* for 1-inch U-Matic, and *critically endangered* for 1-inch SMPTE (Ampex). Here, the structure (record form) becomes a crucial criterion for selection. Context also becomes important because the broadcasting company has a legal deposit

Table 3. Structure, content, and context of some digital objects

	Structure (record form)	Content (information)	Content (linkages)
Videotapes	• VHS, Digital Betacam (Sony), 1-inch SMPTE (Ampex)	• significant national heritage content (e.g. documentaries)	• legal deposit regulation; intellectual property rights of other parties
Web sites	• html, Java applets, etc.	• significant intellectual content	• legal requirements
Digital theses	• word processing software, pdf	• significant intellectual content	• regulations under which submitted; value as a record of the university's activities

obligation for some material it produces, and for other material where it has negotiated rights with intellectual property owners.

For Web sites (the second example) we need to document the structure (record form). This will be html, perhaps with Java applets, CGI, Cold Fusion, and other related applications or add-ons. There is a need to preserve documentation about software. Some Web sites may need to be retained for legal purposes.

For the third example, dissertations and theses in digital form, we need to document the structure (record form) which may be word processing and/or imaging software.

Conclusion

This paper summarizes the criteria traditionally used by libraries when they select for preservation, identifies the main characteristics of the frameworks currently being developed and debated for selecting digital objects for preservation, and compares these with appraisal theory and practice. These summaries are necessarily general — for example, a functional approach to appraisal is changing the basis of appraisal decisions but has not been considered here. Two conclusions can be drawn. First, that new criteria — or, at the least, different weightings of existing criteria — have the potential to assist the library community in advancing its thinking about selection for the preservation of digital objects. And second, that the library community can learn from a closer examination of appraisal theory and practice, in particular the concept of context. These conclusions are supported by Terry Eastwood, who suggests that the cost of preservation of digital objects will become "rather more determinative of the outcome of appraisal" than it was for traditional materials, and that "appraisal is a necessary first step in the preservation process, or so it would seem to judge by the experience of archivists" (Eastwood 2004).

Acknowledgements

Thanks to Richard Cox and Rachel Salmond for providing information and ideas.

Bibliography

Archives Association of Ontario. *What to Keep and What to Destroy?* 2001. aao.fis.utoronto.ca/aa/appraisal.html

Burrows, T. "Preserving the Past, Conceptualising the Future: Research Libraries and Digital Preservation." *Australian Academic and Research Libraries* 31, 2 (2000), pp. 142–153.

Cedars Project. *Working Definitions of Commonly Used Terms (For the Purposes of the Cedars Project).* 1999. www.leeds.ac.uk/cedars/documents/PSW01.htm

Cedars Project Team. *The Cedars Project Report, April 1998–March 2001.* 2002. www.leeds.ac.uk/cedars/OurPublications/cedarsrepmar01exec.html

Cloonan, M.V. "W(h)ither preservation?" *Library Quarterly* 71, 2 (2001), pp. 231–242.

Council on Library and Information Resources (CLIR). Task Force on the Artifact in Library Collections. *The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections.* Washington, DC: CLIR, 2001.

Cox, R.J. "Evidence and Archives." *Records & Information Management Report* 17, 9 (2001), pp. 1–14.

Cox, R.J. *Vandals in the Stacks*. Westport, CT: Greenwood Press, 2002.

Cox, R.J. "No Innocent Deposits." Unpublished manuscript, 2003.

Eastwood, T. "Appraising Digital Material for Preservation as Cultural Heritage" pp. 7–12 in *Preservation of Electronic Records: New Knowledge and Decision-making. Postprints*. Ottawa, ON: Canadian Conservation Institute, 2004.

Edmondson, R. *Memory of the World: General Guidelines*, rev. ed. Paris, France: UNESCO, 2002.

Ellis, J. (ed.) *Keeping Archives*, 2nd ed. Melbourne, Australia: Thorpe in association with the Australian Society of Archivists, 1993.

Gertz, J. "Selection Guidelines for Preservation" in *Joint RLG and NPO Preservation Conference: Guidelines for Digital Imaging*. 1999. www.rlg.org/preserv/joint/gertz.html

Gertz, J. "Selection for Preservation in the Digital Age." *Library Resources & Technical Services* 44, 2 (2000), pp. 97–104.

Gilliland-Swetland, A. *Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment*. Washington, DC: Council on Library and Information Resources, 2000.

Harris, C. "Selection for Preservation." pp. 206–224 in *Preservation: Issues and Planning* (edited by P.N. Banks and R. Pilette). Chicago, IL: American Library Association, 2000.

Howell, A. "Perfect One Day – Digital the Next: Challenges in Preserving Digital Information." *Australian Academic and Research Libraries* 31, 4 (2000), pp. 121–141.

Jones, M., and N. Beagrie. *Preservation Management of Digital Materials: A Handbook*. London, UK: British Library, 2001.

Lyall, J. "Determining the Significance of Documentary Heritage Materials." *Australian Academic and Research Libraries* 24 (1993), pp. 69–77.

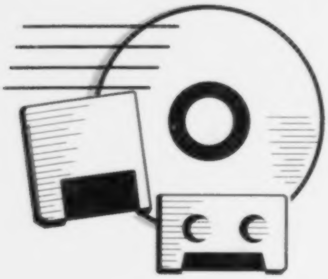
Pacey, A. "Developing Selection Criteria for Special Collections." *Canadian Library Journal* 48, 3 (1991), pp. 187–190.

Russell, R., and K. Winkworth. *Significance: A Guide to Assessing the Significance of Cultural Heritage Objects and Collections*. Canberra, Australia: Heritage Collections Council, 2001.

Taylor, G. *Cultural Selection*. New York, NY: Basic Books, 1996.

Résumé

Les bibliothèques reconnaissent depuis longtemps leur responsabilité en ce qui concerne la préservation des documents pour utilisation future. Les critères appliqués pour la sélection des documents aux fins de préservation, à partir de fonds disponibles surtout sur papier, doivent être réévalués si les bibliothèques veulent s'assurer que les documents sous forme numérique soient aussi préservés. Les critères de sélection qu'elles ont toujours appliqués sont examinés, et cinq critères clés sont établis. On présentera les cadres de référence qui émergent pour la sélection de documents numériques, auxquels on intègre de nouveaux critères de sélection et des facteurs de pondération révisés pour les critères établis. La possibilité d'appliquer des concepts d'évaluation, qui joue un rôle fondamental dans l'archivage, aux pratiques de sélection aux fins de préservation par les bibliothèques est suggérée et mesurée pour certains types d'objets numériques administrés par les bibliothèques.



Selecting for Survival: Developing a Model for Selecting Audiovisual Archival Documents for Preservation Reformatting

Nancy Marrelli

Concordia University Archives
Montreal QC Canada

Abstract

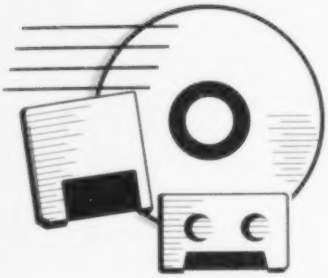
The rapid pace and complexity of ongoing technological change in the highly competitive audiovisual (AV) marketplace presents many problems for the archivist responsible for the preservation of the AV records that document so memorably our society and activities. Costly reformatting of obsolete or endangered formats is essential to ensure ongoing preservation and access, and archivists must partner with technical experts to reformat in the most effective way possible. Alongside this challenging technological maze, selection for preservation is a fundamental part of the process that is often ignored. Archival appraisal and selection for preservation must follow sound archival principles and best practice. This paper proposes the use of an appraisal grid to facilitate responsible decision-making in archival appraisal of AV documents, and presents a model grid that can be adapted to specific institutions and situations to establish priorities for preservation reformatting of AV archival records.

Résumé

Le rythme rapide et la complexité de l'évolution technologique sur le marché très concurrentiel de l'audiovisuel entraînent des problèmes pour les archivistes chargés de préserver les documents audiovisuels témoignant de notre société et de ses activités. Le reformatage coûteux de supports périmés ou menacés de disparition est indispensable à la préservation et à l'accessibilité de ces documents, et les archivistes doivent s'associer à des techniciens spécialisés pour faire ce travail le plus efficacement possible. Outre les difficultés que soulève ce labyrinthe technologique, le choix des documents qu'il y a lieu de préserver est un élément fondamental du processus qui est souvent passé sous silence. L'évaluation archivistique et le choix des documents à préserver doivent respecter de solides principes et pratiques optimales. Nous proposons d'employer une grille d'évaluation qui permettra aux archivistes de prendre des décisions responsables lorsqu'ils évalueront des documents audiovisuels sur le plan archivistique et nous présentons un modèle de grille pouvant être adapté à des institutions et des situations précises afin de fixer les priorités en matière de reformatage de documents audiovisuels archivés pour les préserver.

AUTHENTICITY OF ELECTRONIC RECORDS

AUTHENTICITÉ DES DOCUMENTS ÉLECTRONIQUES



AUTHENTICITY REQUIREMENTS FOR ELECTRONIC RECORDS

LUCIANA DURANTI

School of Library, Archival and Information Studies
University of British Columbia
#831-1956 Main Mall
Vancouver BC V6T 1Z1
Canada
Tel.: (604) 822-2587
E-mail: luciana@interchange.ubc.ca

Abstract

If electronic records are ever to be as trustworthy as records on traditional media, the practices by which they are created, maintained, and used must be analysed, and strategies and standards for their authentic preservation must be developed. This was the mission of a project known as InterPARES (International Research on Permanent Authentic Records in Electronic Systems), which has developed two sets of authenticity requirements for administrative records generated or maintained in databases or document management systems: Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records, and Baseline Requirements for the Production of Authentic Copies of Electronic Records. A second phase of InterPARES has begun studying the applicability of such requirements to interactive, dynamic, and experiential records produced by artistic, scientific, and e-government activities. This paper presents the requirements for authenticity and discusses the challenge posed to their application by the complexity of new types of digital records.

Introduction

In 1922, Sir Hilary Jenkinson identified "authenticity" as one of the characteristics of archival material, and linked it to the procedures of their creation, maintenance, and preservation. Archival documents are trustworthy as statements of fact because those who generate them need to rely on them for action or reference, and they are trustworthy as records because that same need of the creator and its legitimate successor(s) ensures that proper guarantees are put into place to keep them intact over time, both in the short and the long term.[1] With electronic records, however, a presumption of authenticity based on the reliance on them by their creator and on continuous unbroken custody is no longer possible. Because of their manipulabilty, vulnerability, and fragility, the authenticity of electronic records is constantly at risk, especially when they are transmitted across space and time and when they are migrated from an obsolete system to a new one. Thus, authenticity cannot be considered a characteristic of all electronic records, but only of those whose processes of creation, maintenance, and preservation respect certain pre-established requirements aimed at the protection of their authenticity. In all other cases, authenticity must be verified.

Between 1999 and 2001, the InterPARES[2] project identified and articulated requirements that would ensure the protection and permanent preservation of the authenticity of electronic records generated or maintained in databases or document management systems. It did so on the basis of concepts rooted in the traditional disciplines of diplomatics and archival science and elaborated and refined in light of the characteristics of contemporary records. In fact, in order to develop authenticity requirements for electronic records, the InterPARES research team determined that it was necessary to establish what a record was, at least as a hypothesis that the actual investigation could support or overturn. The team decided to adopt the definition of a "record" as any document created (i.e. made or received and set aside for further action or reference) by a physical or juridical person in the course of a practical activity as an instrument and byproduct of it, thereby choosing the traditional archival concept over the traditional diplomatic one, according to which a record is written evidence of a juridical fact. The team then proceeded to define "document" as recorded information, "information" as a message intended for communication across space or time, and "data" as the smallest meaningful piece of information. Finally, an "electronic record" was defined as a record created (i.e. made or received and set aside for action or reference) in electronic form, meaning that a message received in electronic form but set aside for action in paper form is a paper record, while a letter received on paper but scanned in the computer and only used as a digital file is an electronic record. However, the research focused on records born, maintained, and used in electronic form.

Regardless of the choice of an archival definition for an electronic record, diplomatics (the science developed in the 17th century for the purpose of ascertaining the authenticity of records attesting patrimonial and personal rights) was essential to determining the necessary characteristics of such records. The following characteristics were identified:

- a fixed form, meaning that the binary content of the record must be stored so that it remains complete and unaltered, and its message can be rendered with the same documentary form it had when first set aside
- an unchangeable content
- explicit linkages to other records within or outside the digital system, through a classification code or other unique identifier
- an identifiable administrative context
- an author, an addressee, and a writer
- an action, in which the record participates or which the record supports either procedurally or as part of the decision-making process

The other concept that needed to be clarified was that of "authenticity," which was defined as the trustworthiness of the record as a record. In other words, the term refers to the fact that a record is what it purports to be and has not been tampered with or otherwise corrupted. Authenticity differs from authentication, which is one of the means used to prove that a record is authentic at a given moment in time. To understand further what the idea of authenticity implies with regard to its existence and preservation, the research team divided the concept into two components: identity and integrity. Identity refers to the attributes of a record that uniquely characterize it and distinguish it from other records, such as its date and the names of the persons involved in its formation. Integrity is the wholeness and soundness of a record. A record has integrity if it is intact and uncorrupted, i.e. if the message that it is meant to communicate in order to achieve its purpose is unaltered. This means that a record's physical integrity, such as the proper number of bit strings, may be compromised, provided that the articulation of the content and its required elements of form remain the same.

Authenticity requirements

On the basis of these concepts, the team developed two sets of authenticity requirements. The first set, called "Benchmark Requirements," is meant to support the presumption of the authenticity of the electronic records of a creator before they are transferred to the custody of the preserver. The presumption of authenticity is an inference that is drawn from known facts about the manner in which a record has been created and maintained, and is based upon the number of requirements that have been met and the degree to which each has been met. The second set, called "Baseline Requirements," is meant to support the production of authentic copies of electronic records that have been transferred to the custody of the preserver. The two sets of requirements are based respectively on the notions of a trusted record-keeping system and trusted custodianship. A trusted record-keeping system is one that controls what records are included in the system; who can include, retrieve, modify, delete, or remove them from the system; and how the records are included, maintained, retrieved, deleted, or removed from the system. A trusted custodian is a person entrusted with the responsibility of preserving the records — someone who has demonstrated that they have no reason to alter the records entrusted to their care or to allow others to do so, and who is capable of implementing the necessary measures for the physical and intellectual protection of the records.

Benchmark requirements

There are eight Benchmark Requirements: the first identifies the fundamental information about an electronic record that establishes its identity and allows for the demonstration of its integrity; the other seven identify the types of procedural controls over the record's creation, handling, and maintenance that support a presumption of integrity. In contrast to the Baseline Requirements (which are based on archival science), all Benchmark Requirements are derived from the diplomatic body of knowledge.

Benchmark Requirement 1 — The first requirement prescribes that the value of the following attributes[3] are explicitly expressed and inextricably linked to every record. These attributes can be distinguished into categories, the first concerning the identity of records and the second concerning the integrity of records:

A.1.a Identity of the record:

A.1.a.i Names of the persons concurring in the formation of the record, i.e. name of author, writer, originator, and addressee
A.1.a.ii Name of action or matter
A.1.a.iii Date(s) of creation and transmission: chronological date, received date, archival date, transmission date(s)

- *A.1.a.iv* Expression of archival bond
- *A.1.a.v* Indication of attachments

A.1.b Integrity of the record:

- *A.1.b.i* Name of handling office
- *A.1.b.ii* Name of office of primary responsibility
- *A.1.b.iii* Indication of types of annotations added to the record
- *A.1.b.iv* Indication of technical modifications

These attributes may appear as elements of record form or as annotations on the face of the record (e.g. the date, the name of the handling office), but they are more likely to be metadata linked to the record. It is essential that these attributes be inextricably linked to the record. This means that their presence in separate parts of the system, such as the audit trail, is not only unpractical (because the preserver would have to maintain a very large amount of unneeded information in order to keep the specific data related to a record), but it is not helpful in guaranteeing their permanent accessibility in connection with the record and their ongoing existence. The two primary means of linking these attributes to a record are the record profile and the topic map. A record profile is a form inextricably linked to a record, which includes specific fields for the automatic or manual inclusion of data related to the record; it is very common in electronic records management systems. A topic map visually expresses the characteristics of a record and the relationships among them. Whenever a record is removed from the system for external storage, migrated on the occasion of a system upgrade, or transferred to the preserver, the attributes must go with it, remain inextricably linked to it, and be accessible to the user.

Benchmark Requirement 2 — The second requirement regards access privileges. It prescribes that a presumption of authenticity be supported by the fact that the creator has defined and effectively implemented access privileges concerning the creation, modification, annotation, relocation, and destruction of records. The assignment of the authority and capacity to carry out administrative action on the records must therefore be accompanied by the exclusive technical capability to exercise such responsibility. This is usually done by creating, inside the system, tables of users' profiles that provide differentiated kinds of access depending on the users' administrative competence. However, access control can also be exercised by means of external security systems, such as the exclusive assignment of a key to a location. The effective implementation of access privileges consists of monitoring access through the use of audit trails that record each interaction of a user with a record.

Benchmark Requirement 3 — The third requirement prescribes that the creator has established and implemented procedures to prevent, discover, and correct loss or corruption of records. Examples of these procedures include the making of regular backups of records and their attributes, as well as of the entire system; and ensuring that the backup and recovery procedures will guarantee that, in the event of system failure, all complete updates are reflected in the rebuilt files as are any incomplete operations.

Benchmark Requirement 4 — The fourth requirement prescribes that the creator has established and implemented procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change. To counteract media fragility and technological obsolescence, the creator must plan upgrades to the technological infrastructure of its organization, making sure that the ability to retrieve, access, and use records when the upgrades occur is maintained. In addition, the creator must plan procedures of refreshment of the records, moving them from one storage medium to another, and of migration of the records from obsolescent to new technologies.

Benchmark Requirement 5 — The fifth requirement prescribes that the creator has established the documentary forms of records associated with each procedure either according to the requirements of the juridical system or those of the creator. This requirement derives from the fact that the authors of electronic records feel much freer in their compilation than the authors of paper records, and tend to let the technology, rather than administrative procedure, determine the form of the record. An acceptable compromise is to let the documentary form of a record be determined by workflow control technology, where one can connect each step of a procedure to a documentary form. Also, the creator can customize specific applications for the whole organization, so that all e-mails or all spreadsheets of a certain kind, for example, will present the same form. The control on documentary form must go down to the level of record elements, because this is the level at which the authenticity of the record is maintained and can be verified.

Benchmark Requirement 6 — The sixth requirement prescribes that, if authentication is required by the juridical system or the needs of the organization, the creator has established specific rules regarding which records must be authenticated, by whom, and by what means of authentication. This requirement may be met by linking the authentication of specific types of records to the various steps of the administrative procedure, assigning responsibility to a given officer or an office for authenticating either individual or all records, and determining either a method of authentication valid for the entire organization or specific authenticating instruments for specific types of records.

Benchmark Requirement 7 — The seventh requirement prescribes that, if multiple copies of the same record exist, the creator has established procedures that identify which record is authoritative. One of the greatest problems presented by electronic records is the easiness of reproduction. Innumerable copies of each record may exist in an organization, each copy slightly different from the others either because it resides in a different hard drive of a different computer or because it has modifications voluntarily applied to it by various persons in the organization. It is vital for an organization to know which record is the official one, especially because each instance of the same record is inevitably a copy, either of an original or of a draft. In fact, the original record, which in electronic systems is the first complete and effective record either received (if transmitted across space) or saved to a file in the system (if transmitted across time), ceases to exist after being stored for the first time. When recalled, the stored entity is a copy which, in the best of all possible scenarios, is a copy in the form of original, but, in most cases, is simply an imitative copy. Also a draft, while conceptually remaining the sketch or outline of the definitive document, prepared for purposes of correction and meant to be provisional, will only exist as an imitative or simple copy of the draft first stored. Thus, the official copy of each record will have to be subject to strict procedural controls that will serve as a form of authentication, considering that technologically based forms of authentication like the digital signature only serve when records are transmitted across space, as they usually constitute an obstacle to the maintenance of the record to which they are linked. Of course, when the official record is identified, so is the office of primary responsibility for that record, i.e. the office having the formal competence for maintaining the official records that share the same classification and retention period. This will also help to reduce duplication in the organization and designate accountability for the records.

Benchmark Requirement 8 — The eighth and last requirement prescribes that, if there is a transition of records from active status to semi-active and inactive status that involves the removal of records from the electronic system, the creator has established and implemented procedures determining what documentation has to be removed and transferred to the preserver along with the records. This documentation includes all the information necessary to access the records, to establish their identity, and to demonstrate their integrity. If the system generates record profiles, it will be sufficient to ensure that all records are accompanied by their profile. Otherwise, it may be necessary to transfer with the records audit trails, indexes, data directories, and data dictionaries, etc.

These eight requirements are intended to allow the preserver to assess, in the course of the process of appraisal, the authenticity of the electronic records of a creator before they are transferred to archival custody. They are cumulative, in the sense that the strength of the presumption of authenticity is based on the number of satisfied requirements and on the degree to which each individual requirement is satisfied. If the presumption of authenticity is weak, a verification of authenticity is necessary. The verification of authenticity is the act or process of establishing a correspondence between known facts about the record and the various contexts in which it has been created and maintained, and the proposed fact of the record's authenticity; it involves a detailed examination of the records in all their contexts and of reliable information available from other sources (e.g. audit trails, backups, copies preserved elsewhere, textual analysis, etc.).

Baseline requirements

After the records have been presumed or verified authentic and transferred to the custody of the preserver, their authenticity must be maintained by the preserver. To do so, the preserver must produce authentic copies of the records in question, because the production of authentic copies is the only way to ensure the preservation of electronic records. An electronic copy of an authentic electronic record is authentic if attested to be so by the official preserver, but such attestation must be supported by the preserver's ability to demonstrate that it has satisfied all the Baseline Requirements for the production of authentic copies. Only by virtue of this attestation is the copy deemed to conform to the record it reproduces until proof to the contrary is shown. For this reason, the second set of requirements, the Baseline Requirements, directed exclusively to the preserver, must all be implemented at the highest degree.

There are three Baseline Requirements.

Baseline Requirement 1 — The first requirement prescribes that the procedures and system(s) used to transfer records to the archival institution or program, maintain them, and reproduce them embody adequate and effective controls to guarantee the records' identity and integrity, and specifically that:

- unbroken custody of the records is maintained
- security and control procedures are implemented and monitored
- the content of the record remains unchanged after reproduction

As part of the transfer process, the assessment of the authenticity of the records, which had occurred during the appraisal process, should be verified by ensuring that the attributes relating to the records' identity and integrity have been carried forward with the records themselves (Benchmark Requirement 1), along with any relevant documentation (Benchmark Requirement 8). Once the records have been transferred to archival custody, the preserver must establish many of the controls that were described in the Benchmark Requirements, i.e. the preserver must:

- establish access privileges concerning the access, use, and reproduction of the records within the archives, and ensure that they are implemented and monitored
- establish procedures to prevent, discover, and correct loss or corruption of records, as well as procedures to guarantee the continuing identity and integrity of the records against media deterioration and across technological changes
- establish rules determining responsibility for and means of authentication (if authentication is required)

Baseline Requirement 2 — The second requirement prescribes that the activity of reproduction be documented, and that this documentation includes:

- the date of the records' reproduction and the name of the responsible person
- the relationship between the records acquired from the creator and the copies produced by the preserver
- the impact of the reproduction process on their form, content, accessibility, and use
- the details of any copies of records known not to fully and faithfully reproduce the elements expressing the original identity and integrity

The documentation of the reproduction process is essential for the preserver to fulfil the role of trusted custodian of the record, for the user to have access to the history of reproduction (which becomes an integral part of the history of the record), and for future generations to be able to verify the authenticity of the records.

Baseline Requirement 3 — The third requirement prescribes that the archival description of the fonds containing the electronic records includes — in addition to information about the records' juridical-administrative, provenancial, procedural, and documentary contexts — information about changes the electronic records of the creator have undergone since they were first created. It has always been the function, either explicit or implicit, of archival description to authenticate the records and perpetuate their administrative and documentary relationships, but with electronic records this function has become indispensable. In fact, as original electronic records disappear and an interminable chain of non-identical reproductions follows them, in many cases the researchers looking at the last of those reproductions cannot find in them any information regarding provenance, authority, context, or authenticity. The authentication function of archival description is different from that of a certificate of authenticity, because it is a collective attestation of the authenticity of the records and of all their interrelationships as made explicit in the description rather than being simply an attestation of the authenticity of individual records. One might say that, given the mandatory documentation of each reproduction process carried out by the preserver, archival description is superfluous for the purposes of demonstrating the authenticity of the records copies themselves. However, if archival description summarizes the history of all reproductions, it obviates the need to preserve all the documentation of each reproduction and acts as a certificate of authenticity for the archival fonds.

Conclusion

The authenticity requirements have been tested on several aggregations of records and appear to work well in the context of database and document management systems. However, it has became apparent that technological developments are beginning to interfere with the procedures and forms of records creation, that decision-making is increasingly based on records whose creation and form are discretionary, and that great concern is developing about preserving and verifying the authenticity of new record types that are more and more common worldwide. For these reasons, a second phase of InterPARES has undertaken the development of an understanding of interactive and dynamic records resulting from artistic, scientific, and e-government activities, not only in the later phases of their life cycle but from the moment of their creation, in order to develop knowledge and strategies beneficial to both their creators and their preservers. This phase will be concluded by 2007.

Endnotes

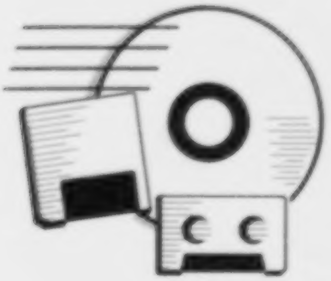
1. Jenkinson, H. *A Manual of Archive Administration.* Oxford, UK: Clarendon Press, 1922, pp. 8–9, 39. Throughout this paper the terms archival document and record will be used interchangeably.

2. The InterPARES (International Research on Permanent Authentic Records in Electronic Systems) project was a collaborative, multidisciplinary, and international research endeavour that involved 60 researchers from 11 countries. Its aim was to develop the theoretical and methodological knowledge necessary for the long-term preservation of the authenticity of electronic records.

3. A record attribute is a defining characteristic of the record or of a record element. A record element is a constituent part of the record's documentary form. An attribute may manifest itself in one or more elements of a record's documentary form (e.g. the name of the author as superscription or as a signature) or in an annotation to the record (e.g. the archival bond as a record identifier) or in metadata in the audit trail, etc.

Résumé

Si les documents électroniques sont appelés à devenir aussi dignes de foi que les documents sur des supports traditionnels, les pratiques permettant de les créer, de les conserver et de les utiliser doivent être analysées, et des stratégies et des normes doivent être élaborées pour leur préservation authentique. C'est la mission qui a été confiée aux responsables d'un projet appelé InterPARES (International research on Permanent Authentic Records in Electronic Systems [recherche internationale sur les documents authentiques permanents dans les systèmes électroniques]) qui ont énoncé deux séries d'exigences en matière d'authenticité pour les documents administratifs générés ou conservés dans des bases de données ou des systèmes de gestion de documents : les exigences-repères à l'appui de la présomption de l'authenticité des documents électroniques, et les exigences de base en vue de la production de copies authentiques de documents électroniques. Une deuxième phase du projet InterPARES a été lancée avec l'étude de l'applicabilité de ces exigences à des documents interactifs, dynamiques et expérientiels produits dans le cadre d'activités artistiques, scientifiques et cybergouvernementales. La communication portera sur les exigences en matière d'authenticité et examinera le défi que pose, pour leur application, la complexité de nouveaux types de documents numériques.



Authenticity and Accessibility, Objects and Technologies: Digital Preservation Basics

Bruce Walton

Library and Archives Canada
395 Wellington Street
Ottawa ON K1A 0N3
Canada
Tel.: (613) 996-1579
E-mail: bwalton@archives.ca

Abstract

Digital preservation is one of the principal challenges facing modern preservation institutions, whether they are archives, libraries, museums, or galleries. A considerable part of the discussion on this subject has concerned the relative merits of different preservation methods, such as emulation and migration. Preservers faced with the choice of a preservation method, while correctly concerned about how best to act, may not in fact appreciate the fundamental nature of the preservation that those methods purport to achieve. This paper therefore discusses the nature of digital materials and what their preservation actually consists of (based mainly on the work of the InterPARES project), and reviews briefly the nature and characteristics of several potential methods as they relate to the twin objectives of digital preservation — accessibility and authenticity.

Introduction

Although action in a given domain is usually best preceded by an understanding of the fundamental nature of the objects and processes involved, this is not always the case. The pressing urgency for action to preserve digital materials in archives, libraries, and museums has led to considerable activity in recent years, one part of which has been a spirited debate regarding the relative merits of specific preservation methods — of which emulation and migration have probably received the most attention.

This is all to the good, but in many cases this activity has proceeded without reference to the fundamental nature of digital objects and digital preservation, presumably because those actively involved in the research and debate have an intuitive understanding of what is involved. Most preservers,[1] however, lack the theoretical and/or practical experience with the subject required for such understanding. The research undertaken in the first phase of the InterPARES[2] project, and particularly that of the project's Preservation Task Force, provides a solid conceptual basis on which preservers may base their work in digital preservation.[3]

Using the InterPARES work as a starting point, this paper considers three questions:[4]

1. What is a digital record?[5] It does not "exist" in the same way as a traditional record — it is not information inscribed on a relatively durable medium and readable without mechanical intervention, so what is it?
2. What exactly is meant by preservation of digital records? For traditional records, it is possible to come up with a basic statement about preservation. Something like "preservation consists of proper containerization, correct storage environments, proper handling when in use, plus copying or stabilization/treatment as required" would probably be accepted by most who are active in the field. But what is the analogous statement for digital records?
3. What methods are available to carry out digital preservation, and how will their use affect the authenticity and accessibility of the records being preserved?

What is a digital record?

To understand digital preservation, it is first necessary to see digital records as existing simultaneously on three linked levels — as conceptual, logical, and physical objects:[6]

- The conceptual object is the record as *human artifact*, able to convey meaning from its creator to its reader, e.g. a textual document, a photo, a map, or a sound recording. It is what we experience with our senses in the everyday world. The conceptual object is what is visible on the computer monitor as a document is created or read/heard/seen; conceptual objects are what get preserved for clients.

- The logical object associated with a given conceptual object is a *computing artifact*, and may be thought of as the conceptual object at the software level; the logical object requires some appropriate software for its use. For the computer, the logical object is what is being worked with as a document is created/edited, saved, or opened.
- The physical object is also a computing artifact, and is as close as one gets to a digital equivalent of a traditional record in storage. The physical object is what the computer is working with when a file manager such as Windows Explorer is used to move or delete a file; the operating system does not need to know whether the file in question is a textual document, a spreadsheet, a photo, etc. — it is simply dealing with a bit string, and no application software is required for the physical object to be acted upon.[7]

There is a fundamental distinction between conceptual and logical/physical objects. The conceptual object is what *humans* recognize and understand; the logical and physical objects associated with a given conceptual object are what *computers and information networks* process, store, and send and receive. In the digital realm (as in the analog), what preservers set out to preserve for long-term accessibility are authentic conceptual objects such as books and records; computers, however, deal only in logical and physical objects. For the computer, the conceptual object doesn't exist — what is seen on the monitor is simply a temporary output or product resulting from the application of some method of reproduction to logical and physical objects.[8] In other words, computers do not and cannot deal with records as records, because records are conceptual in nature, not logical or physical.

What is digital preservation?

The logical and physical object(s) associated with a given conceptual object are what the InterPARES project calls its "digital components."[9] These digital components are all that a preserver has to work with — they are all that there is to "preserve," in the traditional sense of storage and care. The physical maintenance of the digital components over time, while essential, is the lesser part of the challenge of digital preservation,[10] and by itself is not a sufficient action to enable the preserver to present the authentic conceptual object to clients who want it. In order to carry out digital preservation, more than the preserved logical and physical objects are needed — an appropriate method of reproduction is required to enable the conceptual object to be reproduced from its digital components. So the basic statement of what digital preservation consists of, analogous to the basic statement for traditional preservation suggested above, is: *digital preservation consists of maintaining digital components and maintaining the ability to reproduce the record from them.*[11] Put another way, digital preservation requires some combination of digital objects with technology that yields the authentic conceptual object being preserved. Regardless of the method employed to arrive at that objective, the same two functions are being carried out — maintaining digital objects and maintaining the ability to reproduce the authentic record from them.

This definition can be illustrated with an example with which every reader will be familiar — a document attached to an e-mail. Such attachments are opened with one or two clicks of the mouse. The very ease with which the attachment can be opened and displayed might suggest that the attachment *is* the document, but, looked at in the light of the preceding discussion of conceptual objects and their digital components, it be can seen that the attachment is actually the logical object associated with the record, and that the mouse-clicks have called into operation the reproduction method required for that record.

Authenticity and accessibility in digital preservation

The reason it is so easy to read the attachment in the preceding example is that the person using the attached document and its creator inhabit the same technological space (specifically, the user's computer has either the software used by the creator, or another software that can recognize it and reproduce the document). But a technological operation has nonetheless taken place in order to allow the user to see the conceptual object that is the document. The challenge of digital preservation is a dual one. First, how is that reproduction going to work when there are years and decades of technological change between the creation and use of the document, i.e will it remain *accessible*? And second, assuming that there is a working method of reproduction, will the reproduced document be the same in all essential respects to the documents as first consigned to the preserver's custody, i.e. will it remain *authentic*? Any method of digital preservation must keep those things that it is preserving both accessible and authentic.

Accessibility and authenticity are the necessary twin objectives of digital preservation. There is no point in having a completely authentic digital document if it can only be accessed using methods that are technically or financially beyond the means of most preservers, just as ready accessibility is of little value if the authenticity of the thing being accessed is doubtful. Achieving both objectives therefore requires an appropriate balance between the two. As will be seen below, there are methods that, because they involve no changes to the digital components of the document, are very good at maintaining authenticity but, for the same reason, are suspect in terms of maintaining accessibility. Similarly, other methods, because they take care to periodically move the digital components out of obsolete formats into current ones, are good at maintaining accessibility but can threaten authenticity precisely because they change the digital components. So, if digital preservation is about working with digital objects and related technologies in order to produce accessibility and authenticity, how might a preserver actually do this?

What methods are available?

The graphic in Figure 1 shows five different methods that might be used in digital preservation, arranged on a spectrum according to how they address the problem of technology obsolescence.

Imagine a vertical line between "Emulate Old Technology" and "Migrate Formats." The two methods to the left of the line accept technology obsolescence and incorporate it, either wholly or in part, into the "technology" side of the preservation equation. The three methods to the right of the line are focused on avoiding technology obsolescence, or eliminating it altogether. Looking at each method, from left to right:
"Preserve Old Technology" — In this case the digital components are preserved as is, meaning that the only possible method of reproduction is therefore the combination of the application software used to create the record, plus the operating system on which that creating software was designed to run, and the hardware that could run that operating system.
"Emulate Old Technology" — This method is similar to preserving the original technology in that the digital components remain logically as created,[12] but instead of old hardware and software being preserved, they are emulated on new technology, i.e. new software is written to enable a current computer or operating system to behave like the original technology.
"Migrate Formats" — With this method, a line has been crossed in that the digital components are no longer left as is (with the focus of preservation therefore being on maintaining the ability to reproduce the record from the components in their original format); instead, they are changed as required, and the focus of preservation shifts to ensuring that those changes are consistent with ongoing accessibility and authenticity. There are several variants of this approach,[13] but the simplest is *version migration*, in which documents are regularly re-saved in the current version of the software used to create them. Migration is at the mid-point on the spectrum in that, at any given point in time, reproducing the record from its digital components is still reliant on a specific software (as with the two methods already discussed) but, over time, the technology obsolescence problem is avoided (as with the two methods discussed next).
"Standardize Formats" — In this method, the digital components are also changed by being converted to standard non-proprietary formats such as ASCII for databases, tiff for images, ArcInfo Export format for geographic information systems, or pdf or some Web markup language for textual documents.[14] Once in these formats, the records remain accessible because their reproduction is no longer dependent on a specific software, i.e. because the logical object is encoded in a standard form, there are multiple softwares that can be used to reproduce the document.
"Persistent Object Preservation"[15] — This method is still being developed and has been the subject of intensive research in recent years by the United States government [the National Archives and Records Administration (NARA) and other agencies]; initial development of a working system is about to go to contract. This approach to digital preservation, which will be the basis of NARA's

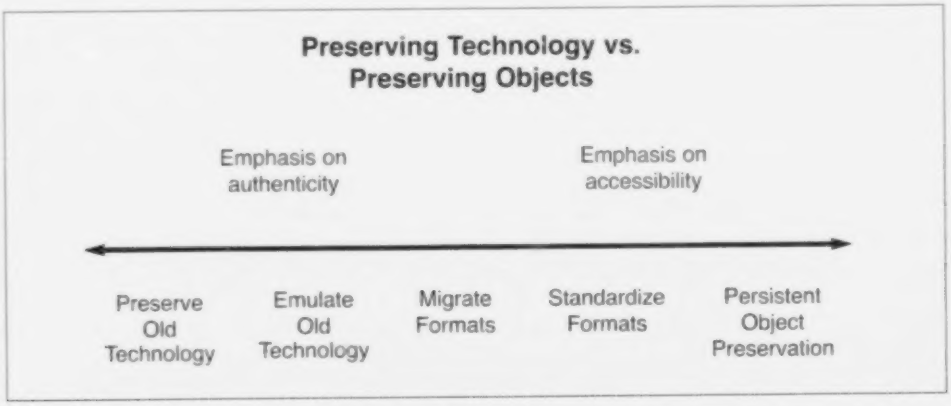


Figure 1. Methods for digital preservation.

Electronic Records Archives,[16] represents a bold attempt to meld record-keeping and archival requirements with advanced computing techniques in order to transcend technology obsolescence. Persistent object preservation is the polar opposite of preserving old technology because it focuses exclusively on the thing to be preserved rather than the technology that will be used to reproduce it. The focus is on all three levels — the conceptual, logical, and physical, the essential characteristics of which are captured in tags. In other words, all aspects of the record are abstracted, or represented by other data, to a very high degree in order to render them as independent of specific technology as possible.[17]

Assessing the options

Looking at the preceding discussion of the five different methods of digital preservation, it can be seen that the two methods that accept old technology (in whole or in part) are focused on the question "how are we going to reproduce an authentic conceptual object from these digital components which we have left in their original software?"; in other words, they are about preserving *technologies*. The other three methods are trying to find the right answer to "how are we going to keep the digital components so that we can use current technology to reproduce the record in an authentic form?"; in other words, they are about preserving *objects*. The methods on the left of the spectrum, by not changing the document, do nothing that might potentially damage authenticity; the methods on the right, by removing the document from its original technological context, avoid the problem of technology obsolescence, and are therefore more likely to preserve accessibility through time.

All these approaches have their pros and cons:

- Preserving old technology offers initial ease of application and no apparent threat to authenticity, but it is simply not practical through time. With ample resources, ingenuity, perseverance, and bit of luck, one or a only a few original technologies might be kept in working order for a decade or two; absent those success factors, and/or with multiple technologies to preserve (which will be the reality in large institutions and, over time, in small ones too), this method is guaranteed to fail — the preserver will be left with unusable digital components. Even if this method could be made to work, it would still not be advisable because records preserved with old technology would require clients to use old technology with which they would be unfamiliar. Nor would it be suited for access over the Internet, where the preserver supplies the client with the digital components and any information necessary to reproduce the document, but the client supplies the technology.
- Emulating old technology, unlike the "preserve" method discussed above, avoids the trap of attempting the impossible task of keeping a museum of functioning technology into the indefinite future; in this sense it is a feasible method. Emulators are sophisticated programs, however, and their development and maintenance is therefore labour- and cost-intensive. This in itself probably means that this method is not suited to most preserving institutions. And even if that were not the case, this method (like the previous one) still requires users to work with old technology and is not suited for access over the Internet.
- Migrating formats overcomes the problem of technology obsolescence by regularly updating the digital components so that they remain accessible with current technologies. This method also has the advantage of being something that most institutions can do now without undue expense. Given the rate of change in the software industry, however, and the fact the backwards compatibility between different versions of the same product is not a high priority for vendors, migration would require ongoing effort to stay on the "software treadmill," which means considerable effort and expense over time. Moreover, the cumulative effect of multiple migrations over time will be to introduce changes to the digital components that could damage the authenticity of the document as reproduced. Finally, quite apart from these considerations, migration's dependence on proprietary softwares puts the entire preservation enterprise at risk — if the product line being used to store the digital components disappears from the marketplace, the preservation effort comes to a dead end.
- Standardizing formats also overcomes technology obsolescence, in this case by converting digital components to widely supported standard formats. It is feasible now (e.g. it is the approach used for archival records at Library and Archives Canada), and allows clients to access records using current technologies — which allows for access over the Internet. On the other hand, conversion to a standard format may be inadequate for some authenticity requirements (e.g. formatted text preserved as an ASCII .txt file would be in a standard form, and therefore accessible, but it would lose many special formatting features of the original). Finally, because standards themselves are subject to change, over time some form of migration — with all that entails — would be required.

- Persistent object preservation, although it has not yet been deployed in an operational system, promises to overcome technological obsolescence and maintain authenticity. One factor to consider is that it has been developed for initial use in an environment (NARA) that must deal with vast collections of records and is sized accordingly in terms of the computing technologies used. Feasibility of use in smaller institutions will depend on this approach being scalable to and affordable in those settings.

As yet there is no consensus in the digital preservation community as to the best method of digital preservation. However, it is worth noting that the four major national archives that have existing digital preservation programs [the National Archives of Canada (NA) and the National Archives and Records Administration (NARA)] or are beginning one [the United Kingdom National Archive (formerly the Public Document Office) (PDO) and the National Archives of Australia (NAA)], all have chosen to use a strategy that emphasizes preserving objects over preserving technologies [conversion to standard format (NA, NARA, NAA) or format migration (PDO)], and that the major research initiative underway among this group (NARA's Electronic Records Archives) also focuses on objects rather than technologies.

Conclusion

This paper has provided a basic overview of the nature of digital records and what is involved in their preservation. It is *far* from being either comprehensive or expert in its coverage of the subject, but may be useful for those faced with digital preservation responsibilities who understand that digital preservation is not the same as traditional preservation but may not be sure how to begin. A little bit of knowledge is a dangerous thing, however, and knowledge uninformed by experience is of less use than knowledge that has been validated and enriched by some real-world practice. Preservers with digital responsibilities should consider the specifics of their situation,[18] identify an appropriate preservation strategy (i.e. select a method), and begin an active digital preservation program, if only on a test basis. In the field of digital preservation (assuming that a copy of the digital components has been set aside just in case), doing something is definitely better than doing nothing.

Endnotes

1. "Preserver" is the term used in the InterPARES research to denote the legal or actual person with responsibility for the maintenance of authenticity and accessibility of digital materials.

2. The InterPARES (International Research on Permanent Authentic Records in Electronic Systems) project was a collaborative, multidisciplinary, and international research endeavour that involved 60 researchers from 11 countries. Its aim was to develop the theoretical and methodological knowledge necessary for the long-term preservation of the authenticity of electronic records.

3. This paper is not formally a report on the findings of the InterPARES project, but its content is very much informed by the results of that project and by the work of Ken Thibodeau (of the United States National Archives and Records Administration) on digital preservation.

4. The physical side of preservation (i.e. using appropriate storage media, storing those media in the correct environment, and periodically copying records forward onto current physical formats) is not considered here. Although physical preservation is a critically important part of digital preservation, it is far more straightforward.

5. Although this paper talks principally of "documents" and "records," the information is generally applicable to electronic publications and other documents in electronic form.

6. See Ken Thibodeau's "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years," which can be found on the Internet at www.clir.org/pubs/abstract/pub107abst.html.

7. The exception to this (in the Microsoft Windows environment) occurs when you try to delete a .exe file — in which case the file manager will ask you if you are sure you want to do this because it could make a program impossible to use.

8. "Method of reproduction" here simply means "software."

9. The full report of the InterPARES Project can be found on the Internet at www.interpares.org/book/index.htm.

10. The preservation of complete bit strings on readable media must be accompanied by whatever changes to the logical format are required by the preservation method being used.

11. See www.interpares.org/book/index.htm.

12. The physical storage format on which they are recorded will of course change over time.

13. See Ken Thibodeau's "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years," which can be found on the Internet at www.clir.org/pubs/abstract/pub107abst.html.

14. It should be noted that pdf is not, strictly speaking, a non-proprietary format. An effort is currently underway to define an archival version (pdf(A)) and to have it adopted as an ISO standard.

15. Given that the method specifically addresses the requirement to preserve structured aggregations of records and not just individual records, the term "Persistent *Archives* Preservation" would be more appropriate.

16. See www.archives.gov/electronic_records_archives/index.html.

17. This approach is explained in far more detail in Reagan Moore's "Collection-based Persistent Digital Archives, Part 1" which can be found on the Internet at www.dlib.org/dlib/march00/moore/03moore_pt1.html.

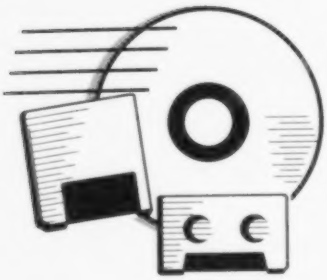
18. That is, preservers should identify the quantity and type(s) of records to be preserved, assess the state of the technology available to them, and consider any specific institutional requirements that apply, such as available budget, in-house information technology and access policies, and relevant legal frameworks.

Résumé

La conservation des données numériques est l'un des principaux problèmes auxquels doivent faire face les établissments modernes voués à la préservation, qu'il s'agisse d'archives, de bibliothèques ou de musées. Une grande part des discussions sur ce sujet ont porté sur les mérites relatifs de diverses méthodes de préservation dont l'émulation et la migration. Les spécialistes devant choisir une méthode de conservation, qui se soucient à juste titre de faire pour le mieux, pourraient ne pas apprécier la nature fondamentale du type de conservation que ces méthodes prétendent offrir. Cette communication examine donc la nature des documents numériques et en quoi consiste véritablement leur conservation (surtout à partir du travail du projet InterPARES), et étudie brièvement la nature et les caractéristiques de quelques méthodes possibles dans la mesure où elles ont un lien avec le double objectif de la conservation de documents numériques : l'accessibilité et l'authenticité.

DEVELOPING A PRESERVATION STRATEGY FOR ELECTRONIC RECORDS

ÉLABORATION D'UNE STRATÉGIE DE PRÉSERVATION DES DOCUMENTS ÉLECTRONIQUES



Is Educating Archivists Enough?: Assessing the Impact of the ACA Institute 2002 "Approaches to the Preservation of Electronic Records"

Amy Marshall

E.P. Taylor Research Library and Archives
Art Gallery of Ontario
317 Dundas Street West
Toronto ON M5T 1G4
Canada
E-mail: amy_marshall@ago.net

Wendy Duff

University of Toronto
Faculty of Information Studies
140 St. George Street
Toronto ON M5S 3G6
Canada
E-mail: duff@fis.utoronto.ca

Abstract

This paper examines the impact of a workshop on electronic records preservation in terms of the progress made by attendees 6 months later. The Association of Canadian Archivists (ACA) Institute 2002 "Approaches to the Preservation of Electronic Records" was a 3-day advanced workshop that brought together archivists from a range of institutions across Canada and expert leaders in the field. A follow-up questionnaire 6 months after the workshop gathered data on how participants viewed the quality of the ACA Institute and how they and their institutions had continued to implement the knowledge gained at the workshop. Despite a low response rate, there was evidence that confidence levels in terms of dealing with electronic records preservation had increased among participants. Participant comments also suggested that there is a need for more concrete education on the topic of electronic records preservation. The authors discuss the success of the ACA Institute and other Canadian educational offerings, and compare alternative models for the delivery of electronic records preservation education.

Introduction: Archival education in Anglophone Canada

Professional education opportunities for Anglophone archivists in Canada fall into two general categories:

- post-graduate academic education (which consists of Master's-level programs offered by three universities: the University of British Columbia, the University of Manitoba, and the University of Toronto)
- post-appointment training in the form of workshops offered by provincial, territorial, and national archival professional associations

There are also some continuing education certificates offered by two colleges in Ontario that fall somewhere in between.

Within the framework of a 2-year Master's program, each of the three universities offers the equivalent of at least one course on managing electronic records that includes a preservation component. Total course hours spent on the problem of electronic records preservation ranges from about 6 elective hours at the University of Toronto to a full elective course dealing entirely with the subject at the University of British Columbia.

At the post-appointment level, educational offerings vary by province. Generally speaking, demand for specialized workshops is low enough that a topic such as electronic records preservation can be offered at most only every few years. In Saskatchewan, for example, a 2-day workshop was offered in 1997, but has not been repeated. Newfoundland and Labrador mounted an introductory workshop on electronic records in 1999. The Toronto Area Archivists' Group, a chapter of the Archives Association of Ontario, offered a 1-day workshop on electronic records preservation in the spring of 2003. Several provinces have no such workshop on record.[1]

The Association of Canadian Archivists (ACA) Institute

The ACA Institute is an annual educational opportunity that addresses timely topics in an advanced forum. Instructors are chosen from international experts in the given topic. Participants are typically archivists with a number of years of experience in the profession, and may be from Canada or other countries. Prior to the Institute, participants are given extensive readings to be reviewed in advance of the workshop. The Institute itself presents theory and debate (much more so than a typical workshop) and often results in new thinking on the given topic. Practical exercises drawn on theoretical content are also included. The Institute is coordinated by the ACA's Education Committee, which has responsibility for archival education and training standards, the provision of continuing education and training opportunities, and support for Canadian archival associations and institutions.

The ACA Institute 2002 "Approaches to the Preservation of Electronic Records"

The ACA Institute 2002 (held in Vancouver, BC, May 27–29, 2002) was the first 3-day workshop to focus solely on electronic records preservation for archivists in Canada. Amy Marshall acted as Institute Coordinator, with responsibility for developing the topic and securing instructors, as well as overseeing the delivery. The topic was chosen for its timeliness and importance to the Canadian archival community.[2]

The 2002 Institute was taught by three leaders: a facilitator, Wendy Duff, who was responsible for leading the group for the full 3 days, encouraging discussion, and serving as a constant point of contact for participants; and two experts, Hans Hofman (Netherlands Ministry of the Interior) and Anne Van Camp (Research Libraries Group), who each provided a 1/2-day perspective on electronic records preservation from their respective vantage points. Each of the leaders also remained present to make comments and answer questions relating to the two Institute assignments. The 2002 Institute also featured a Recorder, John Roberts (New Zealand National Archives), who was contracted to attend the Institute and provide an informal report on its proceedings.

In keeping with past years' Institutes, participants were provided with an extensive reading list in advance of the event. The list was broken into subheadings, and certain readings were prioritized as being "essential" while others were designated as "further reading" on a given topic. A number of digital preservation Web resources were included for browsing purposes. The reading list was presented electronically on the Institute Web site.[3]

Participants

There were 29 participants (28 from Canada and 1 from the Cayman Islands) at the Institute 2002, representing a broad range of archival institutions. Categorized according to type of institution, the breakdown was as follows:

- 3 federal government ministry or department
- 6 provincial government ministry or department
- 6 municipal government
- 6 university
- 4 business
- 1 museum
- 3 other

One municipality and two universities sent multiple representatives.

In the words of John Roberts, Institute Recorder:

> The organisations ranged... in size from the National Archives to sole-charge archives, with a greater concentration of participants from smaller institutions. This may represent a greater reliance on opportunities such as the Association's Institutes for archivists in the smaller organisations to engage with "big" professional issues, where staff of large public archives may be in a position to explore such debate in-house... Participants from all types of archives spoke of the diverse electronic systems — Electronic Document Management (EDM), databases, legacy systems, web applications and many others — which are presently confronting them with preservation and other archives management issues.[4]

Structure

Day 1 of the Institute provided a basic introduction and overview of the state of digital preservation research. Basic preservation principles were covered, and several research programs and strategies were introduced, including InterPARES, emulation and the Universal Virtual Computer, migration, XML, and the OAIS model. The topic of preservation metadata was emphasized. To put some of this new knowledge to work, participants then undertook a brief group exercise consisting of a case study on the preservation of an e-mail document. This case study was designed to build on the presentations of the InterPARES project and the OAIS model and to help participants better understand the attributes of an authentic record and metadata. E-mail preservation provided a manageable, familiar topic.

On Day 2, participants were introduced to the idea of the attributes of a trusted digital repository, including discussion of institutional responsibilities and certification. The basics of the Dutch digital testbed project and the CAMiLEON research project were presented, and the morning concluded with a short overview of policy development by Hans Hofman. In the afternoon, participants proceeded to the major exercise of the Institute — the compilation of electronic records preservation strategies and draft policies. For this exercise, participants were broken into groups based on the kind of institution from which they came (provincial, university, etc.), so as to allow them to collaborate with others with similar situations and concerns. Each group was presented with an institutional scenario that provided a set of typical circumstances in terms of staff size, resources, administrative structure, key types of electronic records, and legislative context,[5] and asked to address eight different types of information in their policy: guiding principles, laws, requirements, preservation methods, standards, staff, possible funding, and requirements for outsourcing and/or collaboration.

The major activity of Day 3 was the group presentations of the policy exercises. Each group explained what they had identified as the major barriers facing their type of institution, and outlined their policies. Hans Hofman and Anne Van Camp made comments, which were followed by group discussion. A brief wrap-up session concluded the Institute.

As part of the course evaluation process at the conclusion of the Institute, participants were asked to consider giving their written permission to be contacted in 6 months' time as part of a follow-up study. Because both authors of the study had been involved in the Institute, we felt that people should have the opportunity to choose whether or not to participate. As it turned out, only 22 participants gave this permission; the other 7 did not wish to be contacted again.

The follow-up study

Survey methodology

The follow-up study involved the administration of a questionnaire to the 22 Institute participants who had agreed to take part. Each participant received:

- an advance-notice letter telling them the questionnaire would arrive 1 week later
- the questionnaire accompanied by a cover letter and a stamped return envelope
- a follow-up letter 1 week after the questionnaire that was a thank-you for their participation and a reminder to return their questionnaires if they had not already done so

Limitations

The chief limitation of the study was the small size of the research population, which ruled out any quantitative analysis of survey results. This limitation was perhaps inevitable, given the nature of the study. A further limitation was the fact that both researchers had been closely involved in the presentation of the Institute, thereby potentially influencing the number of responses and their nature.

Of the 22 people who had agreed to participate, only 15 responded. It is probable that the timing of the questionnaire (late November, which may be a particularly busy or difficult time for many) was a factor in the poor response rate. A follow-up mailing might have increased the number of responses. It is also possible that those who did not respond have made no progress on the problem.

Findings

The 15 completed questionnaires that were received represented a response rate of 68% in terms of the number of questionnaires sent out (n=22), and 54% in terms of the total population of Institute attendees (n=29).

In terms of questionnaires sent out, the responses can be broken down as follows:

Type of institution	No. of questionnaires sent	No. of responses
federal archives or ministry/department	2	1
provincial archives or ministry/department	4	3
university	4	4
municipal	5	3
business	4	3
museum	1	0
other	2	1

Section 1: General reporting and percentages

The first three questions were intended to gauge the perceived quality of the Institute, according to the same criteria used for the course evaluations administered at the conclusion of the Institute. Subjects were asked to rank the following elements of the course on a scale of 1–5, with 1 representing "not satisfied" and 5 representing "very satisfied."

The course material (suggested readings):
33% rated 5
60% rated 4
7% rated 3

The instructors (i.e. the "experts," not including the facilitator):
40% rated 5
47% rated 4
13% rated 3

The content of the course:
20% rated 5
53% rated 4
27% rated 3

The subjects' comments in this section shed light on concerns with the limitations of the course. Several said that the Institute was not hands-on or practical enough. Others wanted more time for the exercises, or wanted readings further in advance and ranked according to priority. Some made positive comments, for example about the policy exercise and the comprehensiveness of the reading list.

Section 2: Individual change

The questionnaire also gathered information about what changes had come about for the individual and his/her institution as a result of the Institute. For example, participants were asked if electronic records preservation was more of a priority for them as a result of attending the Institute, and 13 of the 15 respondents (86%) replied in the affirmative. However, many qualified their answer, saying that preservation had always been a priority. Others suggested that changes within the organization, or in legislation, were the factors responsible for making it more of a priority. One commented: "I used to avoid the issue, now I would like to see some progress on developing a strategy."

Participants were also asked if their confidence level had changed, and if so, in what way. Fourteen of the 15 respondents said it had changed, with all but one indicating that they were more confident. Comments from the respondents revealed that the Institute had raised their awareness of the issues and helped them feel more knowledgeable, but that it had not provided many (if any) solutions. Moreover, a number of respondents suggested that they no longer felt as alone or inadequate because of not having practical solutions. They now had a sense of the magnitude and difficulties involved in digital preservation, and therefore they felt that instead of waiting for a perfect solution they should try to implement some sort of program, as there was no "magic bullet." One respondent commented that s/he now understood the scope of the problem, which s/he found intimidating.

A number of questions were included about how the importance of electronic records preservation issues had changed within participants' institutions. Some respondents suggested that electronic records had indeed become more of a priority in their institutions in the 6 months since the Institute, but that this change had occurred because of legislative or other changes, rather than as a result of the Institute. For a few respondents, prior awareness of the problems of electronic preservation at an institutional level had led to their attendance at the Institute. Four respondents (27%), however, felt that there had been changes in terms of departmental collaboration, and one respondent described the impact as follows: "I think we have been more confident in asserting ourselves in areas where we feel we have responsibilities."

As previously noted, 14 respondents (93%) rated the reading material as either 4 or 5 on the 5-point scale. The respondents' comments about the reading list were generally enthusiastic, with some indicating that they liked having an extensive list. However, others wanted a much briefer bibliography or a list with essential reading highlighted. The bibliography was grouped into essential and other readings, but perhaps the essential list was too extensive for some participants. The questionnaire asked whether participants had continued to consult the reading list, and though respondents had rated the course material highly, only 5 (33%) of the participants said that they had continued to refer to it. A few suggested that they had referred others to it, and others said that they planned to use it in the future but had not yet had time to do so.

Assessing the survey results

The reading list

Reading lists are an essential part of all ACA Institutes. Instructors assume that participants are familiar with the readings, and that having access to readings prior to and following the workshop is essential. However, in the case of a rapidly evolving topic such as digital preservation, the long-term value of the bibliography is questionable. For example, the Dutch Digital Longevity Program Knowledge Bank provides testimony to the growth in the literature: of the 271 dated citations in their bibliography, 126 (47%) were published in 2002 or 2003, and 22 (8%) of the references are marked as New. Furthermore, many citations quickly become out of date as Web resources disappear or become superseded.

A review of the 2002 ACA Institute Reading List in May 2003 revealed no fewer than 11 non-current hyperlinks, or approximately 16% of the list. This problem was partially mitigated by including links to Web sites with bibliographies that are being updated on a regular basis such as the Dutch Digital Longevity Program or the PADI (Preserving Access to Digital Information) Web site, and many references contained in the bibliography (such as the Public Records Office or the InterPARES Web sites) continue to add relevant material related to digital preservation. However, the bibliography does not contain references to relevant documents published in the past year such as the NINCH *Guidelines in Good Practice*, the OCLC/RLG Working Group on Preservation Metadata report *Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects*, and the ERPANET resource database. Neither the instructors nor the ACA are able to update the bibliography, since the Institute was only held once, but the respondents' rating and comments point to its importance.

Evaluation of Institute exercises

As previously noted, many respondents wanted the Institute to include more practical, hands-on exercises. John Roberts' report pointed out that an online e-mail exercise would have allowed "hands-on exploration of the ease (or difficulty) of accessing components of an e-mail other than those typically found in a print-out, of relevant details elsewhere in the environment, and better consideration of the structure of the e-mail message format."[4] For practical reasons, the case study included a single example of a typical e-mail message, rather than an e-mail system or a more realistic quantity of diverse e-mail types existing in most e-mail systems. The case study resulted in a discussion of the intrinsic elements contained in e-mail messages that should be preserved, rather than more technical and/or administrative concerns that may have emerged if participants had been presented with a digital example or a system description.

The reaction to the policy exercise was mixed for a number of reasons. The dynamics of the business archivists' group seemed to impede the progress of drafting a policy, and the group of museum archivists seemed more concerned about the limited financial resources in their archives and overcoming these issues than in drafting a policy. Interestingly, the policy exercise seemed to be most successful for the university archivists' group. Three of the six participants in this group came from the same institution, where they had just begun working on an electronic records policy; and two others in the group worked together. The presence of established collaborators added greatly to the exercise.

Roberts[4] provides an interesting overview of the exercise:

> All groups… confronted almost immediately the relationship of a specific electronic records policy to overall organisational recordkeeping policy. Many of the core principles which groups articulated were of equal importance to recordkeeping in any medium. A widespread view was that an electronic records policy should therefore be developed in conjunction with an overarching policy statement on recordkeeping.
>
> Several groups explored the relationship between the archives and the records management functions. Similarly, questions were raised about the extent to which the archives should be "doing" electronic records itself, or should be working along with business units, and IT units… This also raised questions of the authority for the policy. Does the Archives have the appropriate level of authority for the sort of obligations being imposed by the policy? Will policy be effective if it is seen as a tool of the Archives' or should it be issued by the parent organisation?

Policy on its own will not change institutional practice, and without a widespread awareness of the problems and adequate support, both financial and administrative, a policy will do very little. Furthermore, the participants pointed out that an electronic records policy must be coordinated with other institutional policies and procedures.

Future educational initiatives

Is awareness enough?

As we have discussed, the greatest personal benefits that respondents reported from the Institute were a feeling of not being alone and a greater awareness of the issues. They were appreciative of the networking opportunity provided by the Institute, and felt that the resource of the reading list would be an ongoing benefit.

A major challenge in providing education on electronic records preservation is surely the rapid rate of evolution of the problem. A course curriculum, once developed, is unlikely to be current enough to be offered again a few years later. In particular, bibliographies and resource lists are quickly outdated. The Saskatchewan Council for Archives and Archivists developed such a bibliography for its 1997 workshop. Web-mounted, its hyperlinks were maintained until April 2001, but its content was soon superseded. The duplication of effort on these bibliographies, and the rapid rate at which they are superseded, suggests that resource lists would work better as an ongoing cooperative effort at the international level.

A glance at educational opportunities across Canada suggests that there is inadequate expertise and a shortage of educational opportunities. This shortage is perhaps compounded by the presupposition of many existing education programs that participants bring a level of institutional expertise and resources that may not exist in a majority of smaller archives. A prevailing assumption is that each institution will approach the problem within its own walls, rather than cooperatively with other institutions or associations. So far, the independent approach seems to have had little success.

Alternative models for education

ERPANET is a Europe-based collaborative educational project, fully funded by the EC, which aims to serve as a "virtual clearing house and knowledge base" on digital preservation. Its activities include advisory services, literature review and commentaries, advanced research workshops, and educational opportunities. ERPANET training seminars take a modular approach, offering discrete 2-day learning experiences on such topics as "Policies for Digital Preservation," "The OAIS Model," "Metadata," and "Preserving the Web." Seminars are held in a variety of European locations and feature a variety of international expert speakers. Typically, seminar papers are made freely available on the ERPANET Web site following the seminar. Registration costs are minimal (50 Euros), and seminars and workshops are multilingual when necessary. The ERPANET bibliography takes the form of a database of key resources with analytical reviews. It is available only to registered users, but registration is free.

Cornell University Library has developed an intensive workshop entitled "Digital Preservation Management: Implementing Short-Term Strategies for Long-Term Problems." The structure of the course includes an online tutorial of six modules to be completed by participants in advance of the workshop. The tutorial is also publicly available as a stand-alone educational tool. The workshop, which debuted in August of 2003, is 5 consecutive days in length and is led by one keynote speaker and eight instructors. Registration literature encourages institutions to send delegates in pairs in order to take away the maximum benefit. The workshop is supported by the U.S. National Endowment for the Humanities, which includes subsidization of the registration fee.

Conclusion: Current and future needs

There is a need in the Canadian archival community for more sustained and long-term educational opportunities. The ERPANET and Cornell models are two possible approaches which could be emulated within Canada. An alternative would be to import existing international education programs, or to develop affordable ways to participate in them from afar, such as Web conferencing. However, for any future educational initiative, our survey results suggest that the opportunity to network and to develop a feeling of shared challenges are key social elements of a program. Some important managerial aspects of the electronic records preservation problem, such as legislation, funding models, and institutional culture, have national or regional bases. These are compelling reasons to consider a made-in-Canada approach.

The questionnaire results suggest that bibliographies or resource lists are a highly valued but rapidly outdated element of an electronic records preservation education program. Given the obvious tendency for repetition and redundancy in the core content of bibliographies on this topic, it may be most effective for bibliographies to be maintained on an international, collaborative basis. However, it is important that there be a mechanism for contributing Canadian-specific content to these resources.

Given the small size and overlapping needs of Canadian information professional communities, it may be that archivists need to overcome the limits of traditional professional field boundaries. For example, most electronic records preservation problems extend into the domain of records managers and information technology professionals. The challenges facing archives are shared by similar institutions, such as digital libraries and space data institutions. It would surely be advantageous to share expertise and educational opportunities, particularly for training beyond the introductory level.

To develop any kind of national education strategy on the electronic records preservation problem, it is obvious that substantial funding will be necessary. Both the ERPANET and the Cornell models for education delivery depend upon major government subsidies. Without this kind of support, educational opportunities like the ACA Institute will remain fragmented and infrequent, and it will be difficult to progress beyond elementary course content which is rapidly superseded and reaches few constituents.

Endnotes

1. Canadian archivists can also take advantage of graduate and post-appointment training in a number of related fields and programs. McGill University's Graduate School of Library and Information Studies has recently advertised to hire a digital archives specialist. ARMA International and its regional chapters offer workshops on topics such as managing e-mail, which archivists may opt to attend. The Professional Learning Centre at the Faculty of Information Studies, University of Toronto, offers continuing education courses that include e-mail management and records management, the latter available through distance learning. The Canadian Conservation Institute workshop on digital media covers physical aspects of electronic records preservation.

2. The Institute's promotional statement was as follows:

   > Archivists have been aware of the preservation challenges posed by electronic records for a number of years. We have watched the progress of international research teams tackling important theoretical aspects of electronic records management, and, more recently have seen a number of reports beginning to conclude on aspects of preservation. Yet many archivists remain unsure of how to proceed. There is a need for translation between research findings, their theoretical implications and the best practices for working archivists. The archival community represents diverse institutions with diverse circumstances in terms of resources, expertise, infrastructure, and the receptiveness of parent organizations to support solutions for the preservation of electronic records. The 2002 ACA Institute will aim to address the gap between research and practice by introducing participants to the most important and relevant research findings, and equipping participants to address their own circumstances in terms of policy needs and best practices.
   >
   > Institute participants will gain familiarity, through the perspectives of experts, with the state of the most important international research on electronic records preservation, and will understand the implications of research findings on their own situations. They will come away with the tools to approach electronic records preservation policy-making in their own institutions. The model of the business case study will be adapted to provide archivists with a plausible scenario around which to discuss the perspectives of Institute leaders.

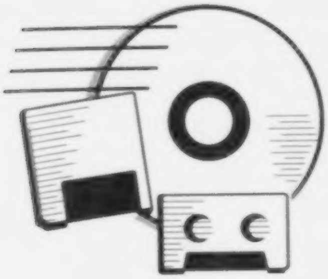
3. The Web site www.archivists.ca/downloads/documentloader.aspx?id=917 features the Institute rationale, leaders' biographies, and the reading list. As in past years, the site was archived following the Institute as a resource for participants and the general public.

4. Roberts, J. "Approaches to the Preservation of Electronic Records: ACA Institute 2002." 2002, unpaginated. Available on the Internet (accessed October 2003) at: www.archivists.ca/downloads/documentloader.aspx?id=915

5. Scenarios were developed by Wendy Duff and Amy Marshall in the winter of 2002 via the administration of a questionnaire to a range of archival institutions, roughly representing the range of institutions expected to be represented at the Institute. Institutions were asked about their resources, holdings (electronic and otherwise), administrative structures, and the factors they perceived as barriers to the systematic preservation of electronic records. Respondents included eight universities, six provincial institutions, four municipal institutions, three corporations, three museums, and one other. Although not scientifically administered, this survey and analysis provided an approximate map of the state of electronic records in Canada. The results were interesting, and the process served as a rough and ready way of developing teaching material for the Institute.

Résumé

Les auteures examineront les répercussions d'un atelier sur la préservation des documents électroniques, plus précisément les progrès réalisés par les participants six mois plus tard. L'atelier organisé par l'Institut de l'Association of Canadian Archivists (ACA) en 2002 sous le thème « Approaches to the Preservation of Electronic Records », d'une durée de trois jours, était une activité de niveau avancé qui a réuni des archivistes de diverses institutions des quatre coins du Canada venus rencontrer des spécialistes de la question. Le questionnaire que les auteures ont distribué par la suite a permis de recueillir des données sur la façon dont les participants ont perçu la qualité de l'Institut et sur les moyens que les participants et les institutions qu'ils représentent ont pris pour continuer à mettre en œuvre les connaissances acquises lors de l'atelier. Malgré un faible taux de réponse, il est ressorti que le degré de confiance est à la hausse parmi les participants en ce qui concerne la préservation des documents électroniques. Les commentaires faits par les participants laissent aussi entendre qu'une formation concrète est nécessaire sur la préservation des documents électroniques. Les auteures traiteront du succès remporté par la formation offerte par l'Institut et d'autres organismes canadiens, et elles compareront d'autres modèles pour la prestation de formation sur la préservation de documents électroniques.



PRESERVING DIGITAL OBJECTS IN SMALL- TO MEDIUM-SIZED INSTITUTIONS

BILL WALKER

Amigos Library Services
14400 Midway Road
Dallas TX 75244-3509
USA
Tel.: (972) 851-8000 ext. 153
E-mail: walker@amigos.org

Abstract

The preservation of digital objects is an issue of pressing concern to all institutions desiring to make electronic resources available to their constituents. While larger institutions strive to meet this challenge, research conducted for Amigos by Technical Data Corporation with funding from the Andrew W. Mellon Foundation reveals that small- to medium-sized institutions are either largely unaware of the problem, unsure of digital preservation strategies and standards, or without the staff and financial resources to adequately confront this challenge. Many of these institutions digitize items of interest to local, regional, and national history. To meet the challenges of preserving digital objects, these institutions need:

- training and other professional development opportunities to inform them of the issues
- standards for preserving digital objects rendered into cost-effective, easily implemented technical solutions
- commitments from their governing entities to fund preservation of their digital objects in perpetuity

Introduction

The preservation of digital objects is an issue of pressing concern to all institutions desiring to make electronic resources available to their constituents. Attempts to deal with the complex issues of preserving digitized and born-digital objects for the indefinite future are taking place all over the world. These include research projects such as Cedars (CURL Exemplars in Digital Objects) in Great Britain, preservation metadata standards produced at the National Library of Australia, and the Metadata Encoding and Transmission Standard (METS) in the United States. International projects and programs such as InterPARES are dealing with the issues of authenticity, reliability, integrity, preservation, and policy. The Open Archival Information System (OAIS) attempts to provide a data model for digital objects entering an archive. Large institutions, governments, and international cooperatives are all committing vast resources to deal with the challenge of digital preservation.

While the larger players try to meet and solve this challenge, small- to medium-sized institutions are unique in that as a group they are either largely unaware of the problem, unsure of how to preserve their digital assets, or without the staff and financial resources to adequately deal with the problem. While larger institutions are more likely to have the funding to send staff to conferences and workshops and invest in leading-edge technological solutions, smaller institutions have little or no knowledge, time, or funding to even begin to understand, much less meet, the challenges of ensuring the authenticity, reliability, and integrity of digital objects. They are unsure whether or not they can guarantee that old files will be readable on new systems.

This paper outlines the challenges facing small- to medium-sized institutions regarding digital preservation and articulates what these institutions need in order to make the task of preserving digital objects for the long term as practical and affordable as current technology will allow.

Background

Amigos Library Services is a library service organization that serves libraries, archives, museums, and cultural heritage institutions throughout the southwestern United States. Amigos began more than 25 years ago as a network providing Online Computer Library Center (OCLC) services, training, and support to its members. It has grown to provide a variety of additional services to more than 800 member institutions. With funding from the National Endowment for the Humanities, the Imaging and Preservation Services (IPS) unit within Amigos provides training, consultation, and support in the areas of traditional preservation and digital imaging. In 1998, Amigos established the digital imaging component of IPS to meet a growing demand for support in digital imaging projects. The Imaging Field Services Officer serves as an information resource and offers a variety of workshops in digital imaging, metadata for digital imaging projects, project management, and preservation of digital objects. Small- to medium-sized institutions send staff to our training workshops, request consultation and site surveys, and call us requesting advice and support.

Amigos classifies its member institutions by type and size, as shown in Table 1.

The challenge

The two main questions before us are:

1. How do we inform small- to medium-sized institutions about the challenges of preserving digital objects?
2. What future solutions will be available to these institutions?

Small- to medium-sized institutions often hold many important documents related to local history, archaeology, and genealogy among other topics. Many of these collections may contain items of interest to local, regional, and national history. For example, one participant at an Arizona workshop represented the Pima Air and Space Museum. This small museum maintains the *International Archive of the Martin B-26 Marauder* (the Martin B-26 Marauder is a lesser-known World War II attack bomber).[1] This is one of the few institutions housing information on this aircraft and its history. Many of these small- to medium-sized institutions have small, often part-time or volunteer staff and look to digitization as a way of making their resources more widely available while protecting the originals from handling by users and staff. For example, the archive of the Museum of International Folk Art in Santa Fe, New Mexico houses the unpublished field notes of anthropologists, art historians, and other scholars who collect folk art from around the world and conduct research in this field.

Statewide and regional initiatives

Practical examples of digital preservation are not yet underway in individual small- to medium-sized institutions, nor have any statewide initiatives dealt with the challenge. The Colorado Digitization Project (CDP) is the most mature statewide initiative in the southwestern United States and has earned a national reputation for its level of cooperation among institutions as well as for the establishment of guidelines for digitization, creation of metadata, and dissemination of knowledge through workshops and other presentations. However, nothing in the CDP Web site or any of its published standards addresses the challenge of preserving digital objects in perpetuity.

Table 1. Amigos classification of member libraries

Type	Size
Academic libraries: 4-year (serving 4-year institutions) Community/Junior Colleges (serving 2-year institutions) Law (serving law schools) Medical (serving medical schools) Theological (serving religious colleges or seminaries) *Public libraries:* Small (in cities with population <50 000) Medium (in cities with population 50 001– 149 999) Large (in cities with population ≥150 000) *Special libraries:* Corporate (serving corporations and organizations, excluding law firms) Law (serving law firms, government law agencies, and other law-related organizations) Medical (serving hospitals and medical research organizations) Museum (serving museums and historical societies)	Small: 1–99 000 volumes Medium: 99 001 – 399 999 volumes Large: 400 000 – 999 999 volumes Research: ≥1 million volumes

Two other states in the Southwest have attempted to follow the model of the CDP. OKDIGITAL is an attempt to gain the participation of cultural heritage institutions throughout Oklahoma in a statewide digitization collaborative. Initially the collaborative plans to gain funding for a project commemorating the Oklahoma Centennial in 2007. However, due to uncertainties surrounding extreme budget reductions, institutions have been hesitant to commit to the project. The New Mexico Digital Collaborative, another attempt to gain statewide participation in digital imaging, has lapsed due to changes in personnel in the state library and other institutions. Neither of these projects has yet digitized any materials or faced the challenge of long-term preservation of digital objects.

Although Amigos Library Services has participated in cooperative traditional preservation efforts such as the Regional Alliance for Preservation, there has been little activity or demand for similar efforts applied to preserving digital objects. Again, the focus is more toward initiating regional digitization projects. The concerns of member institutions drive Amigos' activities, and regional efforts to preserve digital objects will not become a focus of Amigos until a larger number of our members become aware of the issue.

The problem

Training and professional development related to preserving digital materials

Small- to medium-sized institutions urgently need to become aware of the issues surrounding digital preservation. Although these institutions express a need for training in topics related to preservation, emergency preparation, and digital imaging (e.g. scanning, metadata creation, project management), they articulate less of a need for information on the preservation of the objects digital imaging would produce.

Technical Data Corporation (TDC), a consultant serving the non-profit sector, surveyed Amigos member institutions through written instruments, focus groups, and telephone interviews to develop a business plan for IPS. This research was conducted with funding from the Andrew W. Mellon Foundation. One of the survey questions asked respondents to rank existing Amigos workshops from highest interest to lowest interest. Three of the top five workshops dealt with scanning; respondents ranked the digital preservation offering eighth of 17 workshops listed.[2] This prioritization indicates that institutions are more cognizant of their knowledge gap in the creation of digital objects and less aware of their need to preserve them. If these institutions are mindful of the challenges of preserving digital objects over the long term, they do not have the knowledge to make sound decisions regarding the short-term welfare of their files, e.g. refreshing and migration, a challenge with which even the larger institutions grapple. Part of the reason for this ignorance may lie in the fact that smaller cultural institutions often depend on computer services centres operated by their governing agencies to maintain their digital assets. For example, small public libraries often depend on their city governments for their Internet access, Web delivery, and server storage; small academic libraries rely on their institutions' computer services centres. Consequently, these institutions are isolated from the daily rigour of maintaining the network hardware and software and the files served on them, and consequently also the issues related to preserving their digital resources.

These institutions can benefit from easily accessible, cost-effective training that will allow them to quickly become conversant with the issues related to preserving digital objects and to apply this knowledge to meet their local needs. However, as mentioned above, many institutions cannot afford to provide release time or travel funds for staff to attend regional, face-to-face training. Although there are a number of workshops that deal with the issues of preserving digital objects,[3] in many cases the fees for these workshops are out of the reach of many small- to medium-sized institutions. And even when the fees are affordable, travel costs and a lack of release time often prevent staff from participating. To meet the needs of those institutions that cannot pay for travel or release time, Amigos and other similar organizations are offering more of their workshops in self-paced or interactive online versions, which are accessible over the Internet.[4] All of these workshops serve to inform participants of the issues and challenges of digital preservation. They provide strategies for saving files in the short term as well as bringing participants up to date on potential future solutions for saving digital objects for the long term. Nevertheless, to be effective, more of these workshops will need to become available at prices small- to medium-sized institutions can afford.

Not only standards but cost-effective, easily implemented solutions

Much of the focus in digital imaging and the preservation of digital objects has centred upon defining standards for describing, managing, storing, and migrating digital objects. OAIS has defined a data model for how data enter an archive, how they are stored and managed within the archive, and how users retrieve them from the archive.

Projects such as Cedars, NEDLIB (Networked European Deposit Library), and others have clarified some of the issues about what information institutions should record regarding their digital assets. Cedars broad objective was to "...explore digital preservation issues. These range through acquiring digital objects, their long-term retention, sufficient description, and eventual access."[5]

The goals of NEDLIB were to develop a standard system for building a deposit system for electronic publications, define the technical requirements for digital preservation and apply them using emulation, and build a demonstrator system that would apply knowledge gained from the research.[6]

METS provides a standard for transmitting, storing, and archiving the files of a digital object along with its metadata.[7]

All of these standards show potential for meeting the challenges of preserving digital objects. However, the application of all of these standards is still largely uncertain and institutions must engage in a great deal of custom computing in order to apply them. Consequently, these potentially effective ideas will become useful to small- to medium-sized institutions only when they come packaged as economical, easily implemented hardware and software solutions. Implementation should be simple enough that it can be accomplished by staff without extensive technology expertise, or it should be part of the product and service provided by the vendor. Producers of computers, peripherals, and software have already taken sophisticated technologies and made them simple enough to be useful to the most technologically challenged consumer, e.g. scanners, digital cameras, wireless home networks, word processing, and html development software — the list is endless. Technology once available only for those with high-level skills and substantial cash is now available to nearly everyone. The same transformation of technologies for preserving digital objects must take place if small- to medium-sized institutions are to preserve their digital treasures for the long term.

Today, small- to medium-sized institutions benefit from solutions provided by vendors in many of their operations. Shared bibliographic utilities, integrated library systems, reference databases, interlibrary loan systems, and acquisitions services are supplied by vendors in cost-effective, easily implemented hardware and software packages affordable to these institutions. Cooperative ventures among institutions provide high-quality powerful solutions at reduced cost because multiple institutions can pool funds to purchase more than they could individually.

Vendors and membership organizations are beginning to address the problem of preserving digital objects, but the costs are prohibitively expensive even for larger institutions and few, if any, of the products provides a complete solution, particularly with respect to preserving digital objects in perpetuity and overcoming issues of technical obsolescence. OCLC, for example, has launched its Digital Archive, a cooperative where institutions can store digital materials and their preservation metadata in a centralized repository. Aside from its prohibitive annual subscription cost of approximately $12,540, plus a monthly storage cost of $5 per gigabyte, the system accepts a limited variety of image file types. A more cost-effective alternative for small- to medium-sized institutions may be the OCLC Digital Archive Batch Ingest Service, which allows users to submit entire collections of files on electronic storage media to the Digital Archive. The cost is $550 per batch plus the monthly storage charge of $5 per gigabyte. Users submit METS-compliant submission information packages through a program called the Submission Builder. The Digital Archive Batch Ingest Service meets the criteria for easy implementation and lower cost, although in today's economic climate even $550 may be more than many institutions can afford. Regardless of the submission method or cost, the Digital Archive's other drawback is that, while it can check the bit integrity of files (e.g. virus and fixity checks), it cannot yet ensure that files will be readable on future computer platforms.[8] [Note: All costs are in US dollars.]

An alternative to OCLC is DSpace, an open-source software platform that allows institutions to capture, distribute, and preserve digital objects with workflows that follow the OAIS model of submission, archiving, and dissemination. However, as with the OCLC Digital Archive, DSpace has not yet met the challenge of guaranteeing that data will be available in perpetuity. While small- to medium-sized institutions can afford the price of the freeware, the software does not really fit their needs —"DSpace · is organized to accommodate the multidisciplinary and organizational needs of a large institution."[9] The system requires a great deal of resources in terms of hardware and computing expertise to run successfully. Among the skill sets needed to implement DSpace are knowledge of the Unix operating system, JAVA, SQL, and Apache Web Server.[10]

Time is of the essence. Nearly every library in the United States now has an online catalogue based on the MARC transmission standard. As MARC became the established standard for machine-readable bibliographic, authority, and volume-holdings data, making automated library catalogues and circulation modules possible, vendors responded by providing solutions that met the needs of nearly every type and size of library. Libraries could wait until a suitable automated product presented itself because their card catalogues would stand the test of time.

Libraries whose collections exhibit a growing number of digital objects may not have the luxury of time as their digital images become corrupt, unreadable due to obsolescence, or rendered unreliable. Without cost-effective, easily implemented applications, future solutions such as emulation or collection-based persistent digital archives will be of little use to small- to medium-sized institutions.[11]

Funding

Another need of small- to medium-sized institutions is funding. Before institutions can budget for digital preservation, however, cost models upon which to base funding decisions must be available.

One of the questions I ask when presenting workshops in digital imaging is, "Has your institution made the commitment to save your digital assets for the indefinite future?" Without the ongoing commitment of administrators, the survival of an institution's digital assets will be in doubt. The nature of grant funding for digital initiatives has been to pay for digital projects, research projects of a limited scope, and start-up costs for various initiatives. While a grant may fund the cost of a single digital preservation project, the ongoing costs beyond the initial endeavour will be borne by the institution. Therefore, institutions cannot depend on grant funding to preserve their digital objects in perpetuity. It is imperative that their administrations budget for the preservation of digital objects just as they budget for the other activities of the library, archive, or other cultural institution.

Conclusion

Larger institutions have demonstrated the utility of digital imaging as a means of increasing access. These successes have encouraged small- to medium-sized institutions to initiate digitization projects of their own. Increased interest among these institutions for information in digital imaging is evident in the results of the Amigos IPS research. The Business Plan concluded:

> Research revealed that there will continue to be high demand for preservation and especially imaging and digital workshops given the relative lack of knowledge amongst libraries, archives and museums, and the ever-changing technologies and standards of the preservation field.[12]

The Business Plan also stated that museums are in need of training and support on technology issues:

> Key informants from museums expressed that museums may be 'five to seven years behind libraries' when it comes to in-house preservation knowledge and that Amigos IPS could play an important role in educating these organizations. To do so, IPS will need to adapt some of its workshops to the needs of museums.[13]

Following the lead of larger academic institutions, today's public libraries, small and medium community college learning centres, and academic libraries increasingly rely upon electronic books, e-journals, and electronic periodicals. Nearly all of these institutions develop Web pages, which often contain important information of permanent archival value. Institutions accumulate electronic collections without considering the long-term cost and effort that will be required to preserve them. Sometimes these institutions mistakenly view digitization as a form of preservation.[14]

The field of digital preservation is relatively new and clear paths from temporary to long-term preservation of digital assets are not yet apparent even to the larger institutions. Small- to medium-sized institutions suffer from ignorance of the problem or a lack of means to address it.

At the beginning of this paper, I listed two questions. The first ("How do we inform small- to medium-sized institutions about the challenges of preserving digital objects?") can be answered by offering continuing education and other professional development opportunities to staff across the entire spectrum of library and other cultural institutions. Although workshops are becoming increasingly available, these institutions must be able to afford the costs of the offerings in order to be able to participate and workshops should be offered in venues that require a minimum of travel and, thus, staff release time. As research brings solutions for preserving digital objects to fruition, workshops, conferences, and other outlets should inform these institutions of the latest technologies.

In response to the second question ("What types of solutions will be available to small- to medium-sized institutions?"), the preservation of digital objects will only be practical to the institution in question when cost-effective, easily implemented solutions become available. In their present form, Cedars, NEDLIB, OAIS, METS, and the other available standards require custom programming and other technical skills that are beyond the reach of most of these institutions. As research realizes solutions for preserving digital objects, vendors must render these solutions into products that these institutions can afford and implement.

No matter what solutions become available, however, the entities that fund these institutions must be committed to funding the preservation of digital objects for the long term. As it stands now, small- to medium-sized institutions will not be able to rely upon ongoing funding from state or national sources.

Although there are presently no clear answers to these questions, I hope that this paper has raised awareness of the challenges faced by small- to medium-sized institutions in terms of preserving digital objects. One thing we do know is that without affordable and easily accessible training, cost-effective, easily implemented hardware and software solutions, and sufficient funding, the valuable digital assets these institutions hold will be lost forever.

Endnotes

1. For more information on this archive, see Pima Air and Space Museum, *International Archive of the Martin B-26 Marauder*. www.pimaair.org/archives.htm#marauder

2. Technical Data Corporation. *A Business Plan for Imaging and Preservation Services*. Boston, MA: TDC, 2002, p. 14.

3. Conway, P. *Archival Perspectives in Digital Preservation*. www.archivists.org/prof-education/course_catalog.asp#preserv
   Cornell University. *Digital Preservation Management: Effective Short-term Strategies for Long-term Problems*. www.library.cornell.edu/iris/research/workshop.html
   Amigos Library Services. *Preserving Digital Objects in an Uncertain Future*. www.amigos.org/training/courses.html#p

4. Library systems outside of the Amigos region are linking to information on Amigos distance education workshops. See, for examples:
   Central Colorado Library System. *Directory of Online Library Related Courses*. www.cclsweb.org
   Hamilton-Pennell, C. "Getting Ahead Online."*Library Journal* (November 15, 2002). libraryjournal.reviewsnews.com/index.asp?layout=article&articleid=CA256607

5. Cedars Web site. www.leeds.ac.uk/cedars/

6. van der Werf-Davelaar, T. "Long-term Preservation of Electronic Publications: The NEDLIB Project." *D-Lib Magazine* 5, 9 (1999). www.dlib.org/dlib/september99/vanderwerf/09vanderwerf.html

7. Library of Congress. *METS Metadata Encoding & Transmission Standard: Official Web Site*. www.loc.gov/standards/mets/

8. Prices include Amigos Surcharge. Information on product features available from the OCLC Web site. www.oclc.org/digitalarchive/

9. DSpace: Durable Digital Depository: Features. dspace.org/technology/features.html

10. DSpace System Documentation: Installation. dspace.org/technology/system-docs/

11. Moore, R., et al. "Collection-based Persistent Digital Archives: Part 1." *D-Lib Magazine* 6, 3 (March 2000). www.dlib.org/dlib/march00/moore/03moore-pt1.html
    Moore, R., et al. "Collection-based Persistent Digital Archives: Part 2." *D-Lib Magazine* 6, 4 (April 2000). www.dlib.org/dlib/april00/moore/04moore-pt2.html

12. Technical Data Corporation. *Op.cit.*, p. 13.

13. *Ibid.*, p. 14.

14. This has been exhibited by the high number of students enrolling for *Digital Preservation in an Uncertain Future* (an Amigos workshop outlining the challenges and solutions of preserving electronic resources) who expected an overview of basic digital imaging. The description of the workshop is as follows:

    > As libraries increasingly rely on digital resources to provide information to patrons,. as more and more institutions provide digital surrogates to documents, graphics, sound, and video recordings, how will they ensure these resources are available for the long term? Deteriorating storage media, technological obsolescence, and legal issues are all factors that could make today's rich digital resources unavailable tomorrow. This workshop explores the nature of digital resources and hazards that can render them useless, then examines how institutions are combating the problem. Topics include file formats, software, hardware, current strategies like refreshing and migration, possible future strategies, and what you can do to ensure the near-term longevity of your digital resources.

This confusion as to the content of the workshop prompted the change of the workshop title to *Preserving Digital Objects in an Uncertain Future.*

Bibliography

Note: All Web sites were accessed September 22, 2003.

Amigos Library Services. *Preserving Digital Objects in an Uncertain Future* (online workshop description). www.amigos.org/training/courses.html#p

Central Colorado Library System. *Directory of Online Library Related Courses.* www.cclsweb.org

Cedars Web site. www.leeds.ac.uk/cedars/

Conway, P. *Archival Perspectives in Digital Preservation* (workshop description). www.archivists.org/ prof-education/ course_catalog.asp#preserv

Cornell University. *Digital Preservation Management: Effective Short-term Strategies for Long-term Problems* (online advertisement for a workshop and online tutorial to be offered beginning July 2003). www.library.cornell.edu/iris/research/ workshop.html

DSpace: Durable Digital Depository: Features. dspace.org/technology/features.html

DSpace System Documentation: Installation. dspace.org/technology/system-docs/

Hamilton-Pennell, C. "Getting Ahead Online." *Library Journal* (November 15, 2002). libraryjournal.reviewsnews.com/index.asp? layout=article&articleid=CA256607

Library of Congress. *METS Metadata Encoding & Transmission Standard: Official Web Site.* www.loc.gov/standards/mets/

Moore, R., et al. "Collection-based Persistent Digital Archives: Part 1." *D-Lib Magazine* 6, 3 (March 2000). www.dlib.org/dlib/march00/moore/ 03moore-pt1.html

Moore, R., et al. "Collection-based Persistent Digital Archives: Part 2." *D-Lib Magazine* 6, 4 (April 2000). www.dlib.org/dlib/april00/moore/ 04moore-pt2.html

OCLC. *Digital Archive: Storage and Access for Digital Collections.* www.oclc.org/digitalarchive/

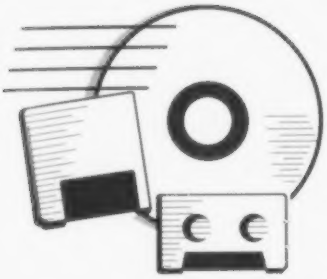
Pima Air and Space Museum. *International Archive of the Martin B-26 Marauder.* www.pimaair.org/archives.htm#marauder

Technical Data Corporation. *A Business Plan for Imaging and Preservation Services.* Boston, MA: TDC, 2002, 31 pp. Unpublished document.

van der Werf-Davelaar, T. "Long-term Preservation of Electronic Publications: The NEDLIB Project." *D-Lib Magazine* 5, 9 (September 1999). www.dlib.org/dlib/september99/ vanderwerf/09vanderwerf.html

Résumé

La préservation des documents numériques constitue une préoccupation pressante pour toutes les institutions désireuses de mettre des ressources électroniques à la disposition de leurs usagers. Les grandes institutions s'efforcent de relever ce défi, mais une recherche menée par Technical Data Corporation pour le compte d'Amigos, grâce à des fonds accordés par l'Andrew W. Mellon Foundation, révèle que les petites et moyennes institutions ne sont pas au courant du problème dans une grande mesure, qu'elles ne sont pas sûres des stratégies et des normes de préservation des documents numériques ou qu'elles n'ont ni le personnel ni les ressources financières nécessaires pour relever adéquatement ce défi. Bon nombre d'entre elles numérisent des documents qui présentent de l'intérêt pour l'histoire locale, régionale et nationale. Si elles veulent réussir à préserver des documents numériques, elles ont besoin de formation et d'autres possibilités de perfectionnement professionnel pour s'informer des enjeux; elles ont besoin que les normes de préservation des documents numériques soient traduites en solutions techniques rentables, faciles à mettre en œuvre; et elles ont besoin que leurs organes directeurs prennent l'engagement de financer la préservation de leurs documents numériques à perpétuité.



DEVELOPING A DIGITAL PRESERVATION STRATEGY AT THE BRITISH LIBRARY: APPLICATION OF THE HANDBOOK *PRESERVATION MANAGEMENT OF DIGITAL MATERIALS*

DEBORAH WOODYARD

Digital Preservation Coordinator
The British Library
96 Euston Road
London NW1 2DB
UK
Tel.: +44 20 7412 7549
Fax: +44 20 7412 7796
E-mail: Deborah.Woodyard@bl.uk

HELEN SHENTON

Head of Collection Care
The British Library
96 Euston Road
London NW1 2DB
UK
Tel.: +44 20 7412 7594
Fax: +44 20 7412 7796
E-mail: Helen.Shenton@bl.uk

Abstract

This paper examines the process of digital preservation strategy development and maintenance at The British Library, and how the handbook *Preservation Management of Digital Materials* has aided in that process. It is a practical example of the application of the handbook's guidance and its translation into real world practice. A summary of the creation of a progress report based on the handbook and actions resulting from this is provided, as well as suggestions for other institutions that plan to develop a digital preservation strategy using this method.

Introduction

Digital information now permeates every aspect of The British Library (BL) collections and business. Every day we are able to provide greater access to collection material through digitization, or use digital publications in new and more efficient ways. Yet we must find a way to stop the digital components of our collections from slipping away as quickly and quietly as they arrived.

Preservation of digital material is a high priority for BL. *New Strategic Directions 2001*, BL's vision for the following 5 years, lists our key responsibilities. They include "ensuring the comprehensive coverage, recording and preservation of the UK national published archive." Because this incorporates a growing proportion of digital materials, BL has developed a dedicated policy and set of strategies for digital preservation. This policy recognizes that preservation cannot be treated as an add-on to other standard library procedures but must be combined into an entire life cycle approach for digital materials. Furthermore, preservation needs to be incorporated in the beginning of that life cycle.

The rapid changes in technology also necessitate constant checks and revisions on BL's work. The policy and strategies for preserving digital collections have been created to be flexible and to support change as it occurs.

Substantial work toward digital preservation began in BL several years ago as we worked toward a digital library infrastructure to manage the long-term access of all our digital collections (Shenton 2000). For example, a digital preservation working group was established in 1999 and the first draft digital preservation policy followed soon after. The policy[1] has been regularly revised for currency and applicability, and in turn it guides the development of relevant and suitable strategies for digital preservation in BL.

A useful tool against which BL's work can be measured is *Preservation Management of Digital Materials: A Handbook* (Jones and Beagrie 2001), first published in 2001 and now maintained by the Digital Preservation Coalition in an online version (www.dpconline.co.uk/handbook.html).

Marc Fresko, director of consulting services at Cornell Affiliates, has also recommended this handbook for explaining the need for digital preservation and providing guidance on the access and management of preserved data. Although he did state that it does not provide a step-by-step instruction manual, he admitted: "But then you won't find it in ready to use form anywhere. Such is the early state of our understanding of digital preservation."[2]

At BL, however, we believed we could devise a process of guiding and cross-checking our work against best practice based on the advice discussed in the handbook. This paper illustrates a practical example of translating the handbook's guidance into real world practice by focusing on the development of BL digital preservation strategies in comparison with the handbook's recommendations.

Libraries and digital preservation

BL's collection of digital material comes in a wide variety of formats and from a number of sources. Since January 2000, BL has received electronic material from publishers under a voluntary code of legal deposit. In addition, BL purchases digital publications, creates digital material through digitization projects, and increasingly collects material from the Internet.

Digital preservation issues bring new managerial and technical challenges into a majority of library areas. The work affected ranges from rights negotiations for acquisition, methods for collection, ability to create catalogue records, etc., through to providing long-term access. It also involves areas normally less directly linked to the collections such as information technology and even corporate records management and marketing sections. Exactly where the responsibilities lie for preservation of digital material is often indistinct, and while this responsibility may be unclaimed by any area of a library it is often partially undertaken by a number of areas operating within the life cycle of the material.

At BL the vast and varied collections meant digital preservation issues were being addressed in several areas with little coordination. Although progress was being achieved, it was desirable to bring a consistent and thorough approach to the efforts by appointing a Digital Preservation Coordinator early in 2001. This provided the necessary focus and dedicated staff time to encourage coherent strategies and ensure BL was working toward best practices. This process included working through the *Preservation Management of Digital Materials* handbook.

Although it can be difficult for smaller organizations to dedicate staff to a similar task, application of the guidance provided in the handbook can help to define where preservation responsibilities occur and enable the allocation of essential roles. It can also provide an authoritative framework for evaluation of all parts of the digital life cycle.

Content of the handbook

The handbook was designed to provide high-level and detailed guidance to institutions, service providers, funding agencies, and creators faced with the challenges of digital preservation in a time where the issues are pressing and little practical experience is available.

It provides a collective set of practices from a variety of sources starting to make headway in the area of digital preservation. It includes BL as a case study and brings together a variety of examples from several sources.

The structure of the handbook includes introductory discussions and a guide to the recommended audience for each chapter as the sections are designed for various levels of knowledg and organizational position. Another useful feature is the set of exemplars provided at the end of each chapter which provide further reference material for future work when engaging in the activities prescribed in each section.

The *Preservation Management of Digital Materials* handbook consists of five sections:

- "Part 1: Introduction" covers the background and use of the handbook and also gives a list of the important definitions and acronyms necessary in discussing digital preservation in this context.
- "Part 2: Digital Preservation" aims to promote understanding of the relevant issues and is particularly aimed at senior administrators, funding agencies, creators, and publishers, or anyone requiring an introduction to the subject.
- "Part 3: Institutional Strategies," "Part 4: Organisational Activities," and "Part 5: Media and Formats" deal with practical issues in detail to help the staff directly involved in this work.

Parts 3, 4, and 5 were used to form the framework fo the progress comparison of BL's digital preservation activities.

It is significant to note that the handbook sections have not been ordered or weighted for prioritization although it is apparent that not all sections would be treated equally within individual institutions. Each institution that considers applying the guidance in the handbook is invited by its authors to use any combination of parts that apply in their own case. After working through the handbook, BL has found that all of the recommended practices are vital to consider and a majority are essential to implement.

Use of the handbook
BL was already heavily engaged in digital preservation work when the handbook was published but value could be seen in using it as the basis for comparing the status of the work already done with currently recommended best practices. This resulted in a progress report for BL and guidance on future strategy development.

The method of comparison used for compiling the progress report is described below, and looking at the report and/or the handbook will provide useful context for the following discussion:

1. Extracting recommended actions from the handbook
Each section of the handbook was combed for recommended practices, which were then laid out in tables. Recommendations were then extracted from the tables and assembled with the same numbering as the handbook sections (this will allow easy cross-checking with the handbook if further detail is required when future strategies are being implemented).

Next, the recommendations were translated into a list of required actions and broken down into elements that specified how the actions could be implemented at BL. Although implementation recommendations were often explicitly listed in the handbook text, some required further interpretation to provide a specific definition (e.g. "Section 4.2.2. Retention and review" — which took extra time and effort).

In several areas (e.g. "External collaboration" and "Rights management"), extra elements were added to cover implementation that had already been carried out at BL that was more detailed and specific than indicated in the handbook.

Not counting these extra BL elements, the first comparative document based on the handbook produced a list of 114 areas of work. This may appear to be a lot but many of these activities already occur in established areas of the digital material life cycle.

This close examination and extraction process led to some observations about using the handbook that might be useful to others.

- To ensure relevance, some sections of the handbook are addressed at more than one level so that the user may choose the most appropriate. For example, "Section 3.5. Staff training and development" sets out advice for institutions, professional associations, and individuals. For this section, BL needed to focus only on the institutional-level recommendations.

- Interpretation of some sections depends on the focus of the user. For example, "Section 3.6. Standards and best practice" endorses the use of standards. But this is not always applicable at BL because, although it is possible to use selected standards for material or metadata created within BL, there is little currently we can affect in the publications that we collect.

- "Section 4.1.2. Creating electronic records" contains a wealth of exemplars and more than 10 categories of further information. Although it is a rich section of the handbook for archive-oriented activities, BL's electronic records are generally corporate records as opposed to collection material and so less important to this progress report. Therefore, BL addressed this section at a high level only for the information of the Electronic Records Management section, and the section was not fully utilized in the comparison exercise.

- The handbook does duplicate some recommendations in different sections. For example, "Section 4.2.3. Accessioning" contains two points about technology and storage that are also covered in "Part 5. Media and Formats." And digital archaeology is discussed in both "Section 4.3.2. Preservation strategies" and "Section 5.4. Technology watch." This duplication within the handbook is necessary to ensure full coverage of a topic in the event that parts or sections of it are used individually.

- It is worth noting that there is an online version of the handbook. This will be kept up to date by the Digital Preservation Coalition in the United Kingdom, and should be consulted (along with other general reading) in addition to the printed handbook. For example, it was noted that one of the preservation research examples in "Section 4.3.2. Preservation strategies" in the printed handbook is already outdated.

2. Evaluation of current status
The implementation elements produced were then evaluated for the current state of progress within BL. Staff in various areas of the digital material life cycle were asked to provide feedback on the elements relevant to their own work. Progress on each item was then rated and added to the tables. Each element was marked as either:

- Not started
- Started
- Ongoing
- Done

to try to simplify the overall view of progress.

3. Analysis of progress

An analysis of progress was made in December 2002 under the assumption that each of the 114 elements identified in the report should be weighted equally. The result was:

- Not started = 46% (53)
- Started/Ongoing/Done = 54% (61)

It was encouraging to see that there had been progress on more than half of the tasks required for digital preservation.

On closer examination of this interesting evaluation, it could be seen that it was not necessarily an accurate representation of the situation. The strategies presented needed to be applied to each different type of digital collection in BL but many had only been addressed in the context of one type of material. Additionally, without any weighting to indicate the priority, the size of task, or the resulting impact of each action, the result was less meaningful. For example, the issues surrounding media formats and physical storage are less significant because there are plans to implement a mass storage system which will eliminate the concern over media format issues.

It was possible to usefully assess a comparative benefit to show that BL had made very good progress in the areas of:

- internal and external collaboration
- outreach
- creating digital surrogates

The areas that required significantly more effort and were a high priority for digital preservation in BL at that time included:

- cataloguing, metadata, and documentation procedures
- technical preservation strategies
- formats and data standards
- technology watch

4. Management acceptance and revision

The progress report that was prepared in December 2002 was prefaced with a summary of the findings and recommendations, and presented to the management team dealing with collections for information and endorsement. The management team agreed that further analysis including weighting of the tasks recommended was required, but generally supported the report for wider dissemination and for the promotion of digital preservation implementation in BL.

5. Dissemination and implementation

The next step was to spread awareness of the progress report to other management and staff and encourage progress on the recommended actions.

Wider dissemination of the strategy was hindered by two issues:

- The length and complexity of the document — With 100 plus elements, the document was too long and detailed for staff members to be expected to read. Staff needed to get past the summary into the detail level, but the recommendations had not been individually addressed for each type of material making it hard for the various collection areas to see what tasks specifically applied to them.
- The lack of specific identification as to where responsibility lays for the ownership of digital preservation strategies — As Varlamoff and Gould (2000) summarized from the synthesis of JISC/NPO studies on the preservation of electronic materials: "What is crucial is that the issue of preservation must be taken into account at all stages, and not just towards the end of the [life] cycle, since the preservation process needs to be considered from the beginning." Because preservation is embedded in every part of the digital life cycle, many areas of BL must be involved. This requires coordination and agreement within the institution on the methods adopted to solve digital material issues; without this kind of cooperation, it is possible that nobody would take on preservation responsibilities, or alternatively that several areas might compete to be the leading driver.

Since the initial progress report at the end of 2002, work at BL has progressed in several areas identified in the report. After some negotiation, many areas in BL are now cooperating to implement digital preservation strategies, working in new and continuing collaborative projects. Significant new activity has especially been occurring in two programs with strong digital preservation components: digital object management and Web archiving.

- Digital object management (DOM) — Work on cataloguing, metadata, and documentation procedures, and to some extent preservation strategies, will be addressed by BL through the development of the planned DOM strategy and systems. In 2003–2004, the DOM program definition phase will introduce an interim storage system for voluntary deposit material and conduct a risk analysis for all digital collection material. After the definition phase, the program will provide the analysis, procurement, and implementation of a full system for long-term digital object management. These large-scale activities will cover many of the recommendations listed in the progress report intended to be applied to the voluntary and legal deposit collections and will possibly also address issues in relation to the Web archive, sound archive, and digitized materials.

- Web archiving — BL is engaging in work on Web archiving. One essential aspect of a preservation strategy for Web sites is the selection and collection of sites from the Internet to ensure they will be kept. Two collaborative projects, one national and one international, will enable BL to address areas such as external cooperation, selection, rights management, metadata creation and capture, formats, storage, access, and long-term preservation strategies for Web material.

BL is also seeking to collaborate with other institutions on a number of preservation research issues, particularly those addressing preservation strategies, formats and data standards, and technology watch as identified in the strategy document. Significant current projects include work with the Digital Preservation Coalition, collaborative research proposals with the National Library of the Netherlands, and participation in working groups such as the Preservation Metadata Implementation Strategies working group (PREMIS).

6. Assessment, revision, and maintenance of the report

While assessing the report for its continuing usefulness as a planning aid, it became apparent that it needed to be refined to be of maximum benefit.

Although the handbook is ideal for listing digital preservation work, it became clear that the complex interdependencies had not been expressed and therefore not carried through into our progress report. Hence our report made no indications as to the priority of recommendations still to be carried out. For example, targets for accessioning tasks cannot be defined until the accessioning tasks are defined; these are dependent on knowing the transfer procedures of materials from producers to BL, which are in turn dependent on rights management and deposit negotiations. These priority judgments need to be made within the institution and applied in planning future work.

Another revision that is required is to apply the strategy to different parts of the digital collections that have different management needs. For example, the rights management issues for the voluntary deposit materials require different and separate work from the digitized images created by BL.

Therefore, to improve the way progress is reviewed for each type of digital material, the report has been reworked. It now consists of a main document of corporate-level work and a series of appendices each addressing a collection that has similar management needs. The six appendices address:

- voluntary and legal deposit materials
- purchased materials
- BL sound archive
- Web archiving
- digitization projects
- document supply material

The report will now provide a more suitable structure for the basis of future strategy planning. Ideally an annual review of progress will be made and interdependencies between recommended actions recognized.

Overall conclusions

Structuring a list of digital preservation work is basically common sense, but using a tool such as the *Preservation Management of Digital Materials* handbook provides a useful and authoritative checklist of an extensive and complex area. This makes it an ideal basis for a review of digital preservation strategies and recommendations for future work.

Based on our experience at BL, we have the following suggestions for other institutions that wish to develop digital preservation strategies by using a comparison of digital preservation work to the handbook:

- The initial evaluation of the handbook might be accelerated by examining the tasks that BL extracted from the handbook, which have been structured in two report framework documents available on the Internet (www.bl.uk/about/collectioncare/digpresintro.html#documents).

- The handbook also exists in an online version (www.dpconline.co.uk/handbook.html). This will be kept up to date by the Digital Preservation Coalition in the United Kingdom, and should be consulted (along with other general reading) in addition to the printed handbook as the implementation elements may change in response to the development of new best practices.

- The handbook does not prescribe a step-by-step solution but does provide enough background knowledge to allow any institution to craft appropriate strategies.

- There is no weighting of priorities or linking of interdependencies in the handbook. This is a task that will need to be done by each institution to suit its own needs.

- Each section may need to be applied repeatedly for different parts of collections if their management needs vary significantly. This can be time consuming but is necessary to provide a consistent strategy.

- A progress report containing the entire list of extractions may be too large to distribute effectively, e.g. the initial report created at BL resulted in an 18-page document. If the report is to be used successfully in strategy development, it would be more productive to create a summary of progress for management, and collection-specific views for other staff.

Endnotes

1. The British Library's digital preservation policy is available on the Internet at: www.bl.uk/about/collectioncare/digpresintro.html#documents

2. Evers, L. "Keeping Up to Date when Preserving the Past." *Network News* 12 (June 2002).

Bibliography

Feeney, M. (editor). *Digital Culture: Maximising the Nation's Investment.* London, UK: The British Library, 1999, 86 pp. www.ukoln.ac.uk/services/elib/papers/other/jisc-npo-dig/

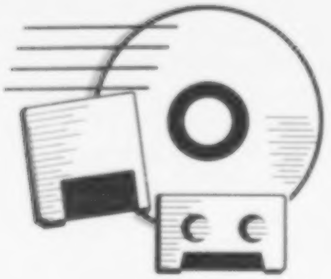
Jones, M., and N. Beagrie. *Preservation Management of Digital Materials: A Handbook.* London, UK: The British Library, 2001, 139 pp. www.dpconline.org/graphics/handbook/index.html

Shenton, H. "From Talking to Doing: Digital Preservation at the British Library." *Preservation 2000 Conference Proceedings.* York, UK: Cedars, 2000. www.rlg.ac.uk/events/pres-2000/shenton.html

Varlamoff, M., and S. Gould. *The Preservation of Digitised Collections: An Overview of Recent Progress and Persistent Challenges Worldwide.* UNESCO, 2001. www.unesco.org/webworld/points_of_views/preservation_1.shtml

Résumé

Nous examinerons le processus d'élaboration et de mise en œuvre de la stratégie de préservation numérique à la British Library et la façon dont le guide intitulé Preservation Management of Digital Materials *a contribué à ce processus. C'est un exemple concret d'application des instructions du guide et de sa traduction dans la pratique réelle. Nous ferons un résumé des conclusions et des mesures issues de cette stratégie et nous offrirons des conseils aux autres institutions qui planifient une stratégie par ce moyen.*



DIGITAL PRESERVATION: BEST PRACTICE FOR MUSEUMS[1]

TIM AU YEUNG

University of Calgary
2500 University Drive North West
Calgary AB T3K 5G2
Canada
Tel.: (403) 220-8975
E-mail: ytau@ucalgary.ca

Abstract

Conservation of collections has long been part of the traditional domain of museums. As creation of digital collections becomes prevalent, museums need to learn about the issues of conserving these new digitally created artifacts that are in fact digital surrogates of objects in their collections. These digital collections range from small, focused collections of digital images created specifically for a CD-ROM or Web-based Virtual Exhibition project to digitization of entire collections. Project planning for digitization should not only encompass the pre-digitization phases, but should also put in place plans for preserving these digital collections once they are created. Various international organizations have looked at issues of digital preservation in the last few years, but few of the papers produced have directly examined the issues from the museum perspective. This paper consolidates the recommendations of past research papers in the context of digital artifacts in the museum environment. It also proposes recommendations that take into account the theoretical issues raised in the past and presents a checklist (or prototype) to allow museums to implement a preservation strategy.

Introduction

Since the advent of computers in the early part of the 20th century, our society has been moving into the electronic world at an increasingly rapid pace. With this movement comes the transformation of our cultural materials from their traditional forms into digital forms, a change that has caused some alarm given the ephemeral nature of the digital environment. In 1994, the Commission on Preservation and Access and the Research Libraries Group launched a task force on digital archiving whose mandate was "to investigate the means of ensuring 'continued access indefinitely into the future of records stored in digital electronic form'" (Waters and Garrett 1996). The need for the task force grew from the realization that our culture and knowledge were in danger of being lost because of the short lifetime of the electronic environment (Beagrie et al. 2001; Hodge 2000). Burgeoning collections of vital cultural artifacts have recently been created, stored, and transmitted in digital form, but these are now housed in cultural institutions ill-equipped to preserve them.

The rapidly evolving nature of the electronic medium (Martin and Coleman 2002) renders technologies obsolete in a remarkably short time, and the ability to access older electronic documents can disappear before our very eyes. As Besser points out, "[t]he artifactual value of electronic art is much different than the artifactual value of more conventional art forms. Because of changing technologies, electronic art originals can only be accessed / viewed / played for a very short time period" (Besser 2000). Libraries in particular have struggled to address these issues as the ease of creating electronic texts has resulted in rapid growth of electronic scholarly publishing. And the problems are rapidly catching up to museums as digitization projects take hold and creators begin to experiment with the unique properties and abilities of the digital medium.

Of particular concern are artifacts that are born-digital and have no physical instance to transcend the digital limitations; born-digital artifacts are those which are "created digitally and [have] functionality which requires use of appropriate hardware and software" (Hanna 2001). To understand a born-digital artifact is to rely heavily on its technological context for meaning (Lee et al. 2002), and saving the digital bits while losing the technical context is still losing the artifact forever. Even the ability to save the digital bits is in doubt as the physical media on which they are stored may become obsolete in as little as 5 years (Rothenberg 1999). And if it is difficult to save one format, consider how the difficulty is compounded when faced with the incredible array of formats that currently exist. This is a real problem for the majority of institutions surveyed in 1998 (Hedstrom and Montgomery 1998) as most had six or more digital formats in their collections at that time. Given the rapid increase in digital collections since 1998, it is not unreasonable to suggest the number of formats has also grown at a comparable pace.

---

[1] This research was commissioned by the Canadian Heritage Information Network.

Spurred by this, an assortment of national and international projects have been trying to find viable solutions to the digital preservation problem. The Victorian Electronic Records Strategy project, PANDORA, Cedars, CAMiLEON, NEDLIB, Kulturarw Heritage, InterPARES, and PRISM (Lee et al. 2002; CLIR 2002) have led the way in exploring digital preservation. But in most cases, their attention has been focused on electronic records and documents rather than on the types of digital objects that museums will encounter in the future. Therefore, we need to look first at the breadth of the literature to understand the state of the art and practice in digital preservation before narrowing our focus to museum-related issues.

The current literature

A point that is continually emphasized in the literature is that creators and publishers must take the burden of digital preservation more seriously (Hodge 2000; Beagrie et al. 2001). Jones and Beagrie note that "[t]he implications of allocating priorities are much more severe than for paper. A digital resource which is not selected for active preservation treatment at an early stage will very likely be lost or unusable in the near future" (Jones and Beagrie 2002). Therefore it makes sense that "creation is where long-term archiving and preservation must start" (Hodge 2000). Creators and publishers, when selecting format and media, need to align their decisions with the support capacities of the institutions that are charged with the task of preserving the objects.

Also consistent in the literature is the need for a distributed approach. One virtue of decentralized preservation is broader cost sharing and redundancy (Reich and Rosenthal 2001), which is very necessary given the enormous scope of digital preservation. Moreover, expertise is unlikely to be concentrated in one institution but instead tends to reside over a broad spectrum of institutions and corporate entities. Indeed, the view of digital preservation as a distributed responsibility is strong enough that the National Library of Australia has enshrined in its statement of principles the idea that "location, selection, identification/cataloguing and retention of digital objects will be best achieved through the coordinated distribution of responsibilities" (NLA 1997). As a result, most projects have emphasized cooperative approaches to digital preservation with interoperability being critical to the solution.

A third area of emphasis is the necessity of preservation metadata. "Effective management of all but the crudest forms of digital preservation is likely to be facilitated by the creation, maintenance, and evolution of detailed metadata in support of the preservation process" (RLG/OCLC 2001). In the case of both the Cedars (Cedars 2001) and the NEDLIB (Lupovici and Masanès 2000) projects, one of the first tasks was to develop a preservation metadata standard to facilitate the rest of the work.

Finally, most digital preservation efforts model their activity on the Open Archival Information System (OAIS) reference model. Initially proposed by the Consultatitve Committee for Space Data Systems for use with space datasets, the reference model has been embraced by the broader digital preservation community as a way that digital archives should work. Both Cedars and NEDLIB have adopted the OAIS model to build their research prototypes. In fact, the joint RLG/OCLC report on the attributes of a trusted repository (RLG/OCLC 2002) goes so far as to make compliance with the OAIS reference model the first attribute of a trusted repository. However, because the OAIS model is a high-level conceptual model, implementation details are not specified; and because off-the-shelf OAIS-compliant systems are not readily available, institutions wishing to comply with the OAIS model must invest heavily to develop a local implementation.

Best practice guides: A typology

After establishing the general issues of digital preservation, we can proceed to developing a typology for comparing and categorizing current best practice documents based on those issues. It is possible to identify eight focus areas addressed in the literature:

1. Conceptualization of digital preservation issues
2. Organizational recommendations
3. Assessment of preservation strategies and recommended methodology

4. Analysis of storage media and digital formats (including life span assessment and recommendations)
5. Metadata standards and practices
6. Resource discovery, persistent identification, and authenticity
7. Intellectual property rights and rights management
8. Cost/resource recommendations and forward-looking statements

As part of the literature survey, a number of guidance and best practice documents were identified (as noted in the Bibliography), and their recommendations categorized using the above typology. It should be noted that this survey was not exhaustive; its goal was simply to identify key documents representing either the advice of organizations leading the digital preservation enterprise or commonly recommended in discussions of digital preservation. It is also important to note that most of the best practice guides are recent and the recommendations may not have empirical support yet. Indeed, the Arts and Humanities Data Service notes a lack of information about how standards and methodology may be applied effectively (Beagrie et al. 2001).

1. Conceptualization of digital preservation issues

Almost all guides identify media deterioration/obsolescence and technology obsolescence as the primary problems associated with digital preservation. In particular, the NINCH guide suggests preservation faces two types of long-term accessibility challenges: machine accessibility (bit integrity) and human accessibility (semantic integrity) (NINCH 2002). Similarly, the Visual Arts Data Service guide distinguishes between physical reliability and continued usability (Grout et al. 2000). There is also general consensus on digital archiving or preservation as a continuous activity in the form of a series of managed activities (RLG/OCLC 2002) or as a life cycle management approach (Jones and Beagrie 2002). On the other hand, the source of digital objects is an area of divergence. A number of guides focus on digitized surrogates of physical objects while others encompass both digitized and born-digital objects with the preservation recommendations biased accordingly.

2. Organizational recommendations

In the area of organizational recommendations, the pressing need for institutional policies on digital preservation is king. As Howard Besser argues, "our community needs to develop a concrete set of guidelines that can be used by people and organizations wishing to make information persist" (Besser 2000). NINCH goes even further to suggest that "[i]nsufficient institutional commitment to long-term preservation can create digital resources with limited sustainability" (NINCH 2002). Beyond the creation of policies, digital preservation needs to be incorporated into the organization as a whole, as in the case of the National Library of Australia where digital preservation is "part of existing core operations and systems … [as opposed to developing] a special or separate undertaking requiring its own infrastructure" (Gatenby 2000). Finally, as noted in the literature survey, most guides reference OAIS as the model for developing a digital preservation architecture.

3. Assessment of preservation strategies and recommended methodology

There is relative uniformity in identifying emulation and migration as the two dominant strategies in digital preservation. However, some guides identify only these two strategies while others present them as two choices in a wider spectrum of possibilities. In most cases, though, there appears to be an implicit acceptance of migration as the primary strategy, ranging from presenting migration as the only strategy (Hodge 2000) to identifying a list of strategies but explicitly recommending migration (NINCH 2002). The Cedars guidance documents are unique in presenting a more nuanced migration with an emphasis on the retention of the original digital object supplemented by migration upon request (Cedars 2002). In contrast, the NINCH guide suggests migrating with every version of a format (NINCH 2002), an undertaking requiring the application of greater resources over the long term. One reason given for migration being favoured is that emulation-based approaches are experimental (Besser 2000); while emulation may be the best hope for complex digital objects in the future, there are few institutions with the technological expertise to create emulators in the short term. In general, there may not be a best strategy but rather efforts should be focused on refining existing strategies (Kenney 2000) with an emphasis on providing a suite of digital preservation tools.

Although there is no clear answer as to which strategy to choose, there is agreement that all digital preservation strategies require that the digital bits be available for future use. To this end, refreshing is identified as the best practice for long-term machine readability. Refreshing can be implemented in a number of ways from immediately moving files onto a common media maintained for all digital objects (Cedars 2002) to more traditional approaches emphasizing environmental stability and routine maintenance and migration of the media (Grout et al. 2000).

4. Analysis of storage media and digital formats

In comparison to strategies, there is less consensus in recommending specific media or formats. This can be attributed to the need to support the significant properties of a broad range of artifacts, often requiring specific formats and media. Even with more general issues such as whether or not to use a compressed format, there is little consensus. While some (NLC 1998) are explicit about holding only non-compressed items, others (Grout et al. 2000) recommend only that the format be lossless to retain maximum fidelity. One argument for compression of any type (including lossless) being problematic for archival files is the introduction of an additional level of complexity (Besser 2000) which only serves to increase the recovery/migration problem in the future. One recommendation that is agreed upon is the use of a standard format that is non-proprietary (NLA 2002; Kenney 2000) as these kinds of formats are more likely to have a preservation path in the future.

One recommendation to handle media issues is to create backups (Cedars 2002; Kenney 2000; IMLS 2001; NINCH 2002) using more than one kind of backup software so as to safeguard against software bugs. In this scenario, at least one copy should be maintained in an off-site location and the media periodically checked as per the refreshing methodology.

5. Metadata standards and practices

There is general recognition of the importance of metadata in an overall strategy for digital preservation (NINCH 2002; Hodge 2000). However, as no single standard for preservation metadata is widely accepted, many organizations and projects (Cedars 2002; Gatenby 2000) have created their own schemes for local use. Fortunately, there are enough common factors between the local schemes that convergence should be possible (RLG/OCLC 2001) with appropriate crosswalks to convert existing metadata to the emerging standards.

For resource discovery metadata and structural metadata, there are emerging standards that have some degree of consensus. Most guides that identify the need for structural metadata recommend the usage of the Metadata Encoding Transmission Standard (METS) (RLG/OCLC 2001; Cedars 2002; NINCH 2002) while Dublin Core is often recommended for resource discovery (Hodge 2000; IMLS 2001; Jones and Beagrie 2002; Grout et al. 2000; Cedars 2002). However, identifying these standards does not imply endorsement and the relative newness of METS (2002) means that implementation guides may not be available for some time.

6. Resource discovery, persistent identification, and authenticity

The need for a persistent identifier to track the object is raised by a number of guides, with proposed systems including PURL (IMLS 2001; Gatenby 2000; Hodge 2000; Kenney 2000), DOI (IMLS 2001; Hodge 2000; Kenney 2000), ISBN/ISSN (IMLS 2001), and local persistent identifiers (Cedars 2002). What is interesting is the surprising number of guides (NINCH 2002; Grout et al. 2000; Jones and Beagrie 2002) that make no mention of the need for a system of persistent identification. Related to the persistent identifier is the need to ensure that the object has not been corrupted or altered. Unfortunately, although the need has been identified, little work has been done except for verification of authenticity at a bit level. For instance, recognition of the difference between authentication (integrity of the record) and authenticity (the quality and context of the record) has been made (Jones and Beagrie 2002), but typically only authentication of the object has been recommended — with the common practice being to calculate a value such as a checksum to ensure bit-level integrity (Kenney 2000; IMLS 2001; Jones and Beagrie 2002). It should be noted that even for this recommendation, the issue of standardization is problematic as checksums can be computed through a number of algorithms and verification requires that the algorithms be identical for the checksum to be useful.

7. Intellectual property rights and rights management

The issue of intellectual property rights and rights management is consistently raised, with the primary emphasis being on ensuring that institutions have sufficient rights for digital preservation activities. Current and proposed changes to copyright laws may prevent preservation either by preventing copying or preventing modifying the digital object (Besser 2000; Gatenby 2000), and this may represent a possible future barrier to preservation. Ideally, an arrangement needs to be made with the copyright holder but, as noted in the experience of the National Library of Australia (Gatenby 2000), the amount of work this represents can be daunting.

8. Cost/resource recommendations and forward-looking statements

Finally, there has not been much discussion in terms of how much these recommendations will cost or even the overall cost of digital preservation. The CAMiLEON project does identify types of costs but the general consensus is that future costs are currently unknown and cannot be well predicted.

The museum context

It is one thing to have a series of high-level recommendations; it is another thing entirely to be able to implement these recommendations. Technology in a museum context can be scarce, and current projects developing digital preservation infrastructures assume a level of technical sophistication not commonly available to museums. A recent survey of technology in museums (IMLS 2002) found that 67% of the museums in the United States have a budget of less than US$250,000 and 20% had no funding for technology. In Canada, the situation is even worse (CHIN 2001) with 60% of the museums having an overall annual budget of less than CAN$100,000 and 85% having a staff of ten or less. To place this in context, one cost estimate (Hendley 1997) suggests that it would take between two and four full-time staff members to manage a digital archive with 100 000 images that was receiving 10 000 new images per year. The obvious conclusion is that any recommendation put before the museum community must restrict itself to action items that are within the constraints of the community.

Recommendations for museums

Based on the overall recommendations provided by the best practice guides and the reality of the environment in which museums operate, two kinds of recommended actions can be provided:

1. Action recommendations (those that each individual museum should be doing immediately)
2. Broader recommendations (those that may require collaboration or an umbrella organization to execute, or may be relevant only in the context of a longer-term view of activities)

1. Action recommendations

- Establish a set of policy documents governing activities related to digital preservation within the institution; a possible checklist of questions to help establish a digital preservation policy is provided in the Appendix at the end of this paper.
- Inventory existing digital holdings and quantify their significant properties; maintain that inventory as the collection grows.
- Assign clear responsibilities for overseeing digital preservation activities to at least one staff member, and enter digital preservation activities into the mainstream operations of the institution.
- Consolidate and reduce the number of media types in the collection and create at least one additional copy for storage in an off-site location.
- Prioritize the relative importance of each format type and the resources allocated to supporting that format; identify formats that the institution will not support and ensure creators/depositors are informed of this.
- Ensure that each digital object in the collection is assigned a persistent identifier with an eye to ensuring that the persistent identifier mechanism is viable beyond the institution.
- Develop a timetable for evaluating holdings including integrity checks of the bit-level data, media refreshing, and retention evaluation.
- Identify a metadata standard that fits with the institution's community of practice, and develop local implementation procedures or adopt an available usage guide to formalize the institutional approach to the usage of that standard.

2. Broader recommendations

- Implement a technology watch protocol to ensure that no media type, file format, or standard becomes obsolete before objects associated with it have been addressed sufficiently.
- Establish collaborative links with other institutions to share expertise and resources; create clusters of expertise in handling particular kinds of digital objects.
- Identify and endorse standards and formats with broad support and sustainable potential, and encourage creators to use those standards and formats for digital objects with enduring value.
- Adopt a system that automates much of the life cycle management of the digital objects.
- Secure from copyright holders rights in perpetuity to copy and modify the object in support of preservation activities.

Conclusion

Given the overall fluidity with which the digital landscape changes, it is doubtful that the techniques of today will be sufficient for the problems of tomorrow. However, the real solution for digital preservation may lie less in technology and more in policy. As Margaret Hedstrom points out, "[t]his challenge is as much a social and institutional problem as it is a technical one, because for long-term preservation, we rely on institutions that go through changes in direction, purpose, management, and funding" (CLIR 2002). For museums, having sound policy that maintains human accessibility to the digital objects is critical. The preservation of digital cultural objects will ultimately be found in the overall commitment to preserve our society's culture and heritage regardless of technical issues.

Appendix: A possible checklist for creating preservation policy

Organizational

1. Selection / acquisition
   a. Is the object important in the context of the institution's core holdings or collection strengths?
   b. Does the object fit into the current or planned digital preservation infrastructure of the institution?
2. Roles and responsibilities
   a. Is the institution the primary holder of record for the digital object?
   b. Has the institution incurred any responsibilities or restrictions for access to the object?
3. Retention / deselection
   a. Are there other institutions with greater capacity or expertise in the type of object at hand?
   b. Does the object fit with the continuing mission of the institution?

Media

1. Choice of media type to use
   a. Does the media type have multi-vendor support for hardware readers and media manufacturing?
   b. Is the media resilient to environmental fluctuations? What are the recommended environmental conditions for long-term preservation and does the institution have the capacity to provide those conditions?
   c. How vulnerable is the media to accidental alteration?
   d. Can the media withstand handling? What are the handling conditions of the media?
2. Management of the media
   a. How long between checks for media readability and integrity? Between media replacement?
   b. Is there an identified off-site location? How often will the off-site store be updated?
   c. Is there an asset tracking system in place for media and how will media be labelled / identified?

Formats

1. Choice of archival format
   a. Does the format have broad support in viewers / editors?
   b. Is the format open / non-proprietary and does it have published specifications?
   c. Does the format support the inclusion of metadata?
   d. Does the format support significant properties of the original (if a digital surrogate)?
   e. Does the format support lossless compression or no compression / encryption?
2. Management of files
   a. Is the version of software that created the file recorded? Is the current version recorded?
   b. How often are format emulators / migrators identified and investigated?
   c. What data loss would constitute a loss of a significant property for the format?

Metadata

1. General
   a. Can the digital objects use a global standard or is there a compelling reason to create a local standard?
2. Resource discovery / descriptive
   a. Does the standard chosen meet the discipline or domain requirements?
   b. Does the standard chosen address interoperability / general resource discovery needs?
3. Structural
   a. Does the standard chosen address the types of aggregation important to the collection?
4. Administrative / preservation
   a. Is there an authenticity indicator (e.g. a checksum) that can be applied to the object?
   b. Can the change history and technological context of the object be traced sufficiently to ensure human readability and authenticity?
5. Persistent identifier
   a. Has the object been assigned an identifier that ensures local uniqueness?
   b. Is it important for the institution to have a universal or global persistent identifier for its objects? If so, which mechanism (e.g. PURL, DOI)?

Intellectual property rights

1. Is the copyright holder's information tracked and stored as part of the metadata?
2. How are the rights of the copyright holder protected from abuse (e.g. limited public access, attribution statement)?
3. Are the usage restrictions consistent with institution policy and mandate?
4. Does the institution have sufficient rights for a preservation regimen?
5. Will the costs of securing rights for long-term access be sustainable over the period of enduring value?

Bibliography

Best practice guides

Note: All Web sites were accessed January 17, 2003.

Besser, H. "Digital Longevity." pp. 164–176 in *Handbook for Digital Projects: A Management Tool for Preservation and Access* (edited by M.K. Sitts). Andover, MA: Northeast Document Conservation Center, 2000. 182 pp. www.nedcc.org/digital/dighome.htm

Cedars Project. *Cedars Guide to Digital Collection Management*. The Cedars Project, 2002, 23 pp. www.leeds.ac.uk/cedars/guideto/collmanagement/

Cedars Project. *Cedars Guide to Digital Preservation Strategies*. The Cedars Project, 2002, 18 pp. www.leeds.ac.uk/cedars/guideto/dpstrategies/

Cedars Project. *Cedars Guide to Intellectual Property Rights*. The Cedars Project, 2002, 19 pp. www.leeds.ac.uk/cedars/guideto/ipr/

Cedars Project. *Cedars Guide to Preservation Metadata*. The Cedars Project, 2002, 18 pp. www.leeds.ac.uk/cedars/guideto/metadata/

Cedars Project. *Cedars Guide to The Distributed Digital Archiving Prototype*. The Cedars Project, 2002, 26 pp. www.leeds.ac.uk/cedars/guideto/cdap/

Gatenby, P. "Digital Archiving — Developing Policy and Best Practice Guidelines at the National Library of Australia." In *Digital Archiving: Bringing Issues and Standards Together*. ICSTI Press, 2000. www.icsti.org/2000workshop/gatenby.html

Grout, C., P. Purdy, and J. Rymer. *Creating Digital Resources for the Visual Arts: Standards and Good Practice*. Visual Arts Data Service, 2000. vads.ahds.ac.uk/guides/creating_guide/contents.html

Hodge, G.M. "Best Practices for Digital Archiving." *D-Lib Magazine* 6,1 (2000). www.dlib.org/dlib/january00/01hodge.html

IMLS (Institute of Museum and Library Services). *A Framework of Guidance for Building Good Digital Collections*. Institute of Museum and Library Services, November 2001. www.imls.gov/pubs/forumframework.htm

Jones, M., and N. Beagrie. *Preservation Management of Digital Materials: A Handbook*. Digital Preservation Coalition, 2002. www.dpconline.org/graphics/handbook/

Kenney, A.R. "Projects to Programs: Developing a Digital Preservation Policy." In *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (edited by A. Kenney and O. Rieger). Mountain View, CA: Research Libraries Group, 2000, 189 pp.

NINCH Working Group on Best Practices. *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*. The National Initiative for a Networked Cultural Heritage, October 2002. www.nyu.edu/its/humanities/ninchguide/index.html

NLC (National Library of Canada Electronic Collections Coordinating Group). *Networked Electronic Publications Policy and Guidelines*. National Library of Canada, 1998. www.nlc-bnc.ca/9/8/index-e.html

RLG/OCLC Working Group on Digital Archive Attributes. *Trusted Digital Repositories: Attributes and Responsibilities*. Mountain View, CA: Research Libraries Group, 2002. 62 pp. www.rlg.org/longterm/repositories.pdf

Museum context

Note: All Web sites were accessed January 17, 2003.

CHIN (Canadian Heritage Information Network). *Information Technology in Canadian Museums: A Survey by the Canadian Heritage Information Network*. Canadian Heritage Information Network, 2001. www.chin.gc.ca/English/Reference_Library/Information_Technology/index.html

Hedstrom, M., and S. Montgomery. *Digital Preservation Needs and Requirements in RLG Member Institutions*. Research Libraries Group, 1998. www.rlg.ac.uk/preserv/digpres.html

IMLS (Institute of Museum and Library Services). *Status of Technology and Digitization in the Nation's Museums and Libraries 2002 Report*. Institute of Museum and Library Services, 2002. www.imls.gov/Reports/TechReports/intro02.htm

Literature surveys

Note: All Web sites were accessed December 18, 2002

Carroll, B.C., and G. Hodge. *Digital Electronic Archiving: The State of the Art and the State of the Practice.* International Council for Scientific and Technical Information: Information Policy Committee, 1999. www.icsti.org/99ga/digarch99_TOCP.pdf

CLIR (Council of Library and Information Resources). *The State of Digital Preservation: An International Perspective.* Council on Libraries and Information Resources, 2002. www.clir.org/pubs/abstract/pub107abst.html

Conway, P. *Preservation in the Digital World.* Council on Libraries and Information Resources, 1996. www.clir.org/pubs/abstract/pub62.html

Fresko, M. "Digital Preservation: State of the Art November 1999 Update." In *Consolidating the European Library Space Conference Proceedings.* DG Information Society, 1999. www.cordis.lu/libraries/events/fp4ce/speech/hs~cox.html

Gould, S., and M.-T. Varlamoff. *The Preservation of Digitised Collections: An Overview of Recent Progress and Persistent Challenges Worldwide.* UNESCO, 2000. www.unesco.org/webworld/points_of_views/preservation_1.shtml

Lee, K.-H., O. Slattery, R. Lu, V. McCrary, and X. Tang. "The State of the Art and Practice in Digital Preservation." *Journal of Research of the National Institute of Standards and Technology* 107, 1 (2002), pp. 93–106.

Waters, D., and J. Garrett. *Preserving Digital Information: Final Report and Recommendations.* Commission on Preservation and Access and The Research Libraries Group, 1996. www.rlg.org/ArchTF/

Other citations

Beagrie, N., D. Greenstein, and C. Pressler. *A Strategic Policy Framework for Creating and Preserving Digital Collections (Version 5.0).* Arts and Humanities Data Service, July 2001.

Cedars Project. *The Cedars Project Report: April 1998 – March 2001.* The Cedars Project, March 2001.

Hanna, M. "Born Digital — Live Digital? Maybe." *ICSTI Forum.* International Council for Scientific and Technical Information, June 2001.

Hendley, T. *Comparison of Methods & Costs of Digital Preservation.* British Library Research and Innovation Centre, 1997.

Lupovici, C., and J. Masanès. *Metadata for the Long Term Preservation of Electronic Documents.* The Koninklijke Bibliotheek, 2000.

Martin, J., and D. Coleman. "The Archive as an Ecosystem." *The Journal of Electronic Publishing.* University of Michigan Press, April 2002.

NLA (National Library of Australia). *Statement of Principles for the Preservation of and Long-term Access to Australian Digital Objects.* National Library of Australia, 1997.

NLA (National Library of Australia). *Safeguarding Australia's Web Resources: Guidelines for Creators and Publishers.* National Library of Australia, June 2002.

Reich, V., and D.S.H. Rosenthal. "LOCKSS: A Permanent Web Publishing and Access System." *D-Lib Magazine.* Corporation for National Research Initiatives (CNRI), June 2001.

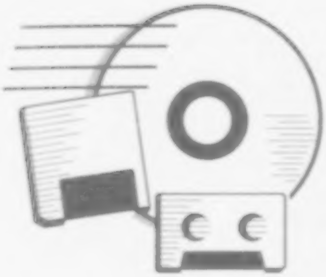
RLG /OCLC Working Group on Preservation Metadata. *Preservation Metadata for Digital Objects: A Review of the State of the Art.* Mountain View, CA: Research Libraries Group, January 2001.

Rothenberg, J. *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation.* Council on Libraries and Information Resources, January 1999.

Résumé

La conservation des collections a longtemps été le domaine traditionnel des musées. À mesure que les collections numériques passent au premier plan, les musées doivent s'initier à la préservation de ces nouveaux objets qui sont, en fait, des substituts numériques des objets de leurs collections. Ces collections numériques vont de la création de petites séries d'images numériques destinées spécifiquement à la production d'un CD-ROM ou à un projet d'exposition virtuelle sur Internet à la numérisation de collections complètes. La planification de la numérisation ne doit pas seulement porter sur les étapes préalables, mais prévoir également la préservation de ces collections numériques une fois qu'elles sont créées. Depuis quelques années, divers organismes internationaux se penchent sur les questions que soulève la préservation des documents numériques. Peu d'études abordent directement ces questions sur le plan muséologique.

Nous examinerons les recommandations issues d'études antérieures concernant les objets numériques des musées. Nous proposerons des mesures tenant compte des questions théoriques déjà abordées et nous proposerons une liste de contrôle ou un prototype permettant aux musées de mettre en œuvre une stratégie de préservation.



NET ART ET CONSERVATION, QUELLES STRATÉGIES POUR LE MUSÉE?

Anne Laforet

Étudiante-chercheur
Laboratoire Culture et Communication
Université d'Avignon
France
Courriel: anne@sakasama.nett

Résumé

Apparu avec le déploiement d'Internet, le Net Art attire depuis quelques années l'attention de musées qui l'exposent, le commandent, l'achètent. La question de l'élaboration par le musée d'une stratégie de conservation spécifique au Net Art se pose alors, non sans de nombreuses difficultés, et ce, malgré l'expérience de l'institution muséale dans la préservation de l'art contemporain qui tend de plus en plus, en partie du moins, vers l'éphéméralité. Les œuvres de Net Art sont fluctuantes par définition et dépendantes de nombreux paramètres pour être vues. Différentes approches sont en train d'être explorées, notamment celle du réseau des Medias Variables autour du Musée Guggenheim de New York et de la Fondation Daniel Langlois à Montréal, approche où l'œuvre n'est plus définie par son médium, envisagée plutôt comme une partition que comme un objet fini. Une démarche mixte qui allierait cette approche muséale avec l'archivage automatique et continu d'Internet peut être envisagée pour poursuivre la réflexion autour des stratégies de préservation.

Le Net Art, création artistique spécifique à Internet, dynamique et multiforme, suscite depuis quelques années l'intérêt des institutions artistiques et muséales, d'ailleurs présentes sur le réseau par le biais de sites Web. Cet intérêt peut surprendre a priori car ces œuvres permettent un rapport sans médiation institutionnelle entre l'artiste et le visiteur de son site. De plus, le musée est une institution basée sur une collection de biens culturels mobiliers inaliénables et le Net Art, une production artistique immatérielle, éphémère, basée sur de l'information et non pas sur des objets. Cependant, les musées exposent, commandent, collectionnent, conservent des œuvres de Net Art.

De manière parallèle au développement d'Internet comme moyen de communication, la création contemporaine qui lui est propre, le Net Art, s'est développée Les possibilités du réseau (informations décentralisées, fonctionnement rhizomique et non pas hiérarchique, etc) ont été exploitées pour créer un art sans médiation entre l'artiste et l'internaute, le créateur et le spectateur/regardeur/visiteur, etc. Un art éphémère, basé sur l'information et non pas autour de l'objet. Un art en perpétuelle mutation du fait du flux de données d'Internet, des changements de protocoles ou de serveurs, des évolutions technologiques. Un art où la notion de paternité explose. Une pratique artistique où les artistes sont leurs propres critiques, galeries, agents, commissaires d'exposition…

L'art spécifique à Internet concerne les œuvres créées par, avec, pour Internet, et plus particulièrement le World Wide Web (ou Web). Il ne s'agit donc pas de toute œuvre d'art présente sous une forme ou une autre sur le réseau, telles des photos d'œuvres « matérielles » numérisées. Ces œuvres d'art Internet sont consultables à tout moment par tout ordinateur connecté au réseau, il suffit de connaître l'adresse de leur serveur. Elles sont plus ou moins ouvertes à l'intervention du visiteur, utilisent des liens hypertextuels, peuvent intégrer du texte, de l'image, animée ou non, du son, des bases de données, etc. Elles peuvent être également liées à des installations ou à des dispositifs dans un lieu « physique ». Le sujet de ces œuvres est très varié. Le matériel informatique nécessaire pour produire les œuvres n'est pas forcément très sophistiqué, et à un prix abordable, ce qui permet aux artistes une certaine indépendance, notamment par rapport aux institutions culturelles et muséales.

Les termes pour définir cette création sont multiples : art Internet, Net Art, net.art, œuvres en ligne, œuvres d'art spécifique au Web, cyberart, art visuel, art électronique, art cybertronique, etc. La création contemporaine par et pour Internet ne s'actualise pas uniquement sous la forme de sites Web, le Web n'étant que la partie la plus « conviviale » d'Internet. Le site Web est ce qui est le plus proche de l'objet, même lorsqu'il est basé sur les flux de données et qu'il n'est pas « signé » par quelqu'un d'identifiable en tant qu' « artiste ». C'est la manifestation la plus évidente et permanente de la création, au détriment d'autres utilisations plus proches de la performance ou des collaborations en temps réel (émission de données en continu ou *streaming*, *chat*, etc). La difficulté de conserver ces pratiques où leur documentation, contrairement au Land Art par exemple, ne fait pas partie de l'œuvre, semble a priori renforcer la préférence des institutions pour le Net Art actualisé en site Web.

La conservation implique des questionnements techniques (matériel, logiciel, supports de stockage, etc), mais aussi juridiques (paternité de l'œuvre, droits d'auteur…), ainsi que la politique de collection elle-même (choix d'œuvres selon la capacité du musée à les préserver, et non pas selon des critères esthétiques, etc).

La conservation suppose préservation et transmission. Conserver, c'est choisir de garder quelque chose d'hier ou d'aujourd'hui pour qu'il soit montré demain. C'est préserver un objet pour qu'il reste présent dans notre mémoire collective.

Il peut sembler paradoxal de vouloir matériellement conserver des œuvres procédant de l'éphémère et de la transformation. « Pourquoi vouloir à tout prix conserver l'art électronique? Ne peut-on envisager un art temporel, qui meurt comme les humains (et en l'occurrence également les machines) qui l'ont porté? Un art sans mémoire? » (1) Néanmoins, il est important de garder au moins des traces des œuvres, ne serait-ce que pour donner des repères au public et aux artistes (par exemple afin que ces derniers n'explorent pas les mêmes idées continuellement).

Les possibilités techniques de stockage de l'information, grâce aux technologies numériques, peuvent donner l'illusion de la préservation, à court terme, de la totalité des productions humaines. La tendance va d'ailleurs vers un traitement automatique de l'information, sans l'intervention d'une personne spécialisée. Toutefois, la conservation induit une sélection, un prélèvement selon des critères précis, définis, tendant à l'objectivité.

Cette sélection s'inscrit dans le temps au sens où elle n'est pas (nécessairement) contemporaine de la création de l'œuvre. En général, une période de temps s'impose pour déterminer si la conservation d'une œuvre est pertinente ou non. L'art contemporain, qu'il soit ou non numérique, a raccourci le temps entre création et conservation.

« Conserver n'est pas une fin en soi. Tout conserver revient à ne rien préserver. La conservation de l'art électronique doit s'appuyer sur les réponses à des questions préalables : que garder? comment prélever? qui choisit? - avant même que de penser aux solutions techniques à mettre en œuvre. » (2)

Ces critères de choix doivent être élaborés par le musée. Ses ressources, autant financières qu'en personnel et matériel, étant limitées, acquérir et conserver l'art spécifique au Web (et faire des recherches dans le domaine) signifie un choix institutionnel autant qu'artistique.

Mettre en place une stratégie de conservation spécifique n'est pas chose facile pour les institutions muséales car le Net Art lui pose des problèmes inédits, malgré l'expérience récente du musée dans la préservation d'autres œuvres éphémères d'art contemporain telles celles de Land Art, de performance, ou d'art conceptuel. Deux domaines d'interrogation se combinent plutôt, d'un côté des problèmes anciens liés à la conservation des œuvres d'art qui resurgissent, et on s'aperçoit alors qu'ils n'ont été résolus que par à-coups ou consensus et qu'ils nécessitent d'être réévalués, et de l'autre, les questionnements spécifiques à la préservation des documents électroniques. L'alliance des deux en fait un champ d'expérimentation à la fois excitant et périlleux.

La fragilité des supports numériques permet difficilement de prévoir l'évolution des œuvres car elle dépend de nombreux facteurs artistiques, technologiques (3) et économiques aux intérêts souvent divergents. De plus, Internet étant un médium fluctuant, la durée de vie des œuvres peut être très courte, selon l'évolution des technologies, de manière assez proche d'une œuvre de littérature orale, et nécessitant un suivi technique régulier.

Afin de pouvoir « maintenir » et « préserver » une œuvre de Net Art, il est d'abord nécessaire de déterminer les éléments qui la définissent. Est-ce le code source de la page Web en language HTML ou bien est-ce l'expérience de l'œuvre pour le regardeur du site? À ceux partisans de la conservation du seul code source, Jon Ippolito, conservateur en arts médiatiques au Solomon R. Guggenheim Museum à New York et artiste, répondait en 1998 « le problème, bien sûr, est que les caractéristiques qui seront probablement perdues dans une telle conversion [du langage HTML] seront exactement celles dont les artistes du Web dépendent actuellement. La plupart des projets artistiques qui sont aujourd'hui faits pour le Web ne sont pas uniquement basés sur une collection spécifique d'informations verbales ou quantitatives, mais sur un design visuel soigneusement contrôlé ou sur un jeu autour du médium qui les encadre. Qu'adviendra-t-il d'une image Bitmap de précisément 800 X 600 pixels quand la résolution d'écran atteindra jusqu'à 10 000 000 X 10 000 000? Cela dégénérera soit en une image minuscule au milieu d'un écran autrement vide, ou bien en une grande image avec une résolution épouvantablement basse. Ce n'est pas une question théorique pour les musées qui prennent en considération l'acquisition de projets Web pour leurs collections. » (4)

Les paramètres qui déterminent la façon dont une œuvre de Net Art sera vue sont nombreux : la version du navigateur et des modules externes, la présence de programmes exécutés côté client et/ou côté serveur, la dépendance du site regardé au réseau (liens vers d'autres sites, accès à des bases de données en ligne, à des caméras Web, à des données sonores ou visuelles en émission continue, etc), le débit de la connexion à Internet, les capacités de l'ordinateur (puissance de calcul, système d'exploitation, propriétés de l'écran, c'est-à-dire résolution, restitution des couleurs, rémanence)… Tous évoluent rapidement et doivent être pris en considération lorsqu'il s'agit de préserver une œuvre. L'artiste ne peut donc pas contrôler la manière dont son travail sera perçu, loin d'une installation interactive dans un espace d'exposition par exemple (où l'artiste est souvent associé à la mise en espace de son travail). Cette diversité de perceptions possibles explique pourquoi certains artistes mettent en place plusieurs versions de leurs sites selon la version du navigateur, le système d'exploitation, la connexion du regardeur.

Au-delà des questions techniques, une question à la fois conceptuelle et pratique se pose au musée : comment délimiter une œuvre de Net Art? Une pièce de Net Art peut fort bien soit être en évolution permanente, soit intégrer des contenus en ligne produits par d'autres personnes, et ce, de différentes manières, par exemple afficher tels quels d'autres sites Web ou bien récupérer des données en flux, etc. Il n'est pas toujours évident, même de la part des artistes eux-mêmes, de percevoir la frontière entre ce qu'ils créent et leur environnement. Cette question est également majeure au plan légal, en ce qui concerne les codes de la propriété intellectuelle. Cette « délimitation » de l'œuvre ne peut avoir lieu que s'il y a connaissance du contexte original de création, qui doit être également préservé autant que possible, car les stratégies de préservation et de présentation sont liées. Par exemple, si une institution faisait l'acquisition de *Shredder* de Marc Napier (5), une célèbre œuvre de Net Art créée en 1998 qui agit tel un navigateur « déchiquetant » le contenu de pages Web choisies par l'internaute (en interprétant le code source de la page de manière différente d'un navigateur traditionnel), devrait-elle la présenter avec les sites internet et la technologie originels ou bien avec les outils disponibles au moment de ses actualisations? Les deux sont possibles, mais portent des significations différentes.

Pour répondre à ces interrogations, Jon Ippolito est à l'origine d'une approche originale, celle de « médias variables » (6) où l'œuvre n'est plus définie par son médium afin qu'elle puisse évoluer, être recréée, par exemple lorsque son médium d'origine sera devenu obsolète. Chaque œuvre est envisagée individuellement, plus comme une partition que comme un objet fini, immuable. Au moment de son acquisition, sa variabilité est prise en compte (pour les sites Web, elle est non intentionnelle et est le résultat des conditions dans laquelle l'œuvre a été créée, conditions qui ne seront à terme plus disponibles) et l'institution dialogue avec l'artiste afin de connaître au mieux son intention, les caractéristiques de l'œuvre, si celle-ci a une « date d'expiration ».

L'approche des médias variables ne concerne pas uniquement le Net Art, mais également toutes les formes d'art contemporain basées sur le processus, et non plus l'objet, tel l'art conceptuel, le Land Art, l'art minimal, les installations... Les comportements de chaque œuvre sont définis à partir d'une grille d'analyse : œuvre installée, performée, interactive, reproduite, dupliquée, encodée, en réseau, contenue. Ainsi, pour revenir à notre exemple, *Shredder* est une pièce à la fois dupliquée, interactive, encodée et en réseau. Dupliquée car sa copie se fait sans perte de qualité. Interactive car le visiteur du site peut manipuler le dispositif. Encodée car écrite en code machine. En réseau parce que consultable sur Internet.

À partir de ces caractéristiques, l'artiste choisit une ou plusieurs stratégies de conservation parmi les quatre identifiées par la réflexion autour des médias variables : le stockage tel quel des données, la migration d'un support à l'autre, l'émulation, la ré-interprétation. L'apport principal du paradigme des médias variables est l'identification d'une quatrième stratégie, la ré-interprétation, pour se dégager de l'aspect physique, technologique de l'œuvre.

Le stockage est la solution la plus classique, qui consiste à entreposer les œuvres sur support numérique. L'œuvre disparaîtra quand ses matériaux ou données seront devenus obsolètes.

La migration implique la mise à niveau d'un support à l'autre. Il y a migration quand on convertit dans un nouveau format un fichier informatique, ou qu'on l'enregistre dans une version plus récente. La migration peut avoir pour conséquence le changement d'apparence d'une pièce, par exemple si des fonctionnalités disparaissent d'une version à l'autre d'un logiciel.

L'émulation consiste à recréer l'apparence de l'œuvre (avec un code source différent). Conserver les ordinateurs sur lesquels ont été créées les pièces n'est pas envisageable à long terme, mais les émuler logiciellement est possible. En installant les différentes couches logicielles qui ont été conservées, il est possible d'exécuter les fichiers, originaux ou modifiés, afin de recréer l'œuvre. Cela semble plus simple et moins onéreux que la migration par exemple car il ne faut plus traiter chaque fichier mais plutôt chaque plateforme informatique, ce qui peut être développé à l'échelle d'un réseau de structures patrimoniales. L'émulation est surtout conseillée pour des programmes autonomes et la question de l'émulation du réseau se pose pour le Net Art. Il ne s'agit pas uniquement d'émuler le débit de connexion mais les protocoles mêmes d'Internet, les programmes côté serveur et/ou client, ainsi qu'également le contenu du réseau selon les cas, comme dans l'exemple de *Shredder*. Malgré les développements qui auront lieu dans les technologies de l'émulation, il est probable que seuls certains éléments d'un travail artistique en ligne pourront être préservés. Ces fragments n'en demeureront pas moins très précieux pour les musées.

La ré-interprétation, stratégie la plus radicale, consiste à re-interpréter l'œuvre à chaque actualisation, à recréer une œuvre fidèle à l'intention de l'artiste (grâce aux données recueillies par le musée suite aux entretiens et questionnaires) mais qui peut s'avérer très différente de l'original. C'est la plus « risquée » mais celle aussi qui permet au musée de changer de rôle par rapport aux œuvres. Pour reprendre les mots de Jon Ippolito, « aussi excentrique que l'idée puisse paraître aux pratiques traditionnelles de collection, [cette vision de la conservation] offre une alternative pour ceux dont la conception de leur œuvre va au-delà de sa manifestation dans une forme particulière. Et elle nous aide à imaginer le musée comme un incubateur pour des œuvres d'art vivantes, changeantes, plutôt que comme un mausolée pour des œuvres mortes. » (7)

Cette attitude par rapport à la préservation de la création artistique contemporaine a été adoptée par plusieurs institutions nord-américaines qui forment actuellement le réseau des médias variables. À la suite du Musée Guggenheim à New York et de la Fondation Langlois à Montréal, sont également membres du réseau le Berkeley Art Museum/Pacific Film Archive, Franklin Furnace Archive, Performance Art Festival + Archives, Rhizome.org et le Walker Art Center.

Je souhaiterais amener en conclusion une réflexion autour d'un modèle de préservation qui allie une démarche muséale, même radicale comme celle des médias variables, avec l'archivage automatique et continuel du Net. L'environnement des œuvres de Net Art (listes de discussion, sites critiques, liens commentés, etc) en effet n'est pas conservé par les musées qui se concentrent sur une œuvre d'art, faisant malgré eux disparaître à terme le contexte original de la création.

Au contraire, un archivage automatique de l'Internet prend en compte l'aspect inter-relationnel de l'œuvre de Net Art qui évolue dans un environnement dynamique. Une indexation régulière et automatisée (ne nécessitant pas l'intervention d'une personne chargée de la conservation) permet de suivre au plus près l'évolution des pièces. L'enregistrement automatique n'implique cependant pas que les œuvres soient en « bon état », semblables à ce qu'elles étaient à l'instant de la captation.

Dans une étude en 1996 consacrée aux arts électroniques, Annick Bureaud, Nathalie Lafforgue et Joël Boutteville avançaient comme modèles de lieux consacrés aux arts technologiques le musée archéologique et la médiathèque. Le musée archéologique semble une option à développer : « Le musée d'art qui accueille des objets uniques n'est certainement plus le modèle de la conservation de l'art électronique, même s'il peut encore remplir marginalement cette fonction ou ouvrir ses espaces physiques à d'autres formes de conservation. Le musée archéologique, en revanche, apparaît comme un modèle plus pertinent : il mêle culture savante et objets du quotidien ; il conserve des « morceaux brisés » (équivalent d'œuvres qui ne « fonctionnent » plus comme elles le devraient) qu'il sait « faire parler »; il travaille sur la répétition et l'accumulation d'objets identiques dans divers états qui permettent une reconstitution mentale de l'état ou des états originels. Cette archéologie est déjà nécessaire dans l'art électronique. En effet, nombre d'œuvres des pionniers ne fonctionnent plus ou mal, nécessitent une remise en état, quand elles ne commencent pas à disparaître purement et simplement. » (8)

Le musée serait donc, dans cette optique, une archive vivante, un espace de documentation, avec des fragments d'œuvres qui peuvent être réactualisés de multiples façons. Cela permettrait d'aborder la création en ligne en combinant une approche quantitative (automatisation de la création d'objets informationnels) avec une démarche plus qualitative, celle d'un conservateur d'œuvre d'art.

Notes :

(1) dans Boutteville, Jöel, Bureaud, Annick et Nathalie Lafforgue. *Art et technologie : la monstration* (en ligne), Délégation aux arts plastiques (DAP), Ministère de la Culture et de la Communication, 1996, 200 pages environ, section IV.II.1., mise en ligne été 1998. [http://www.olats.org/livresetudes/etudes/monstration/section4ok.shtml#grand2]

(2) idem. section IV.II.1. [http://www.olats.org/livresetudes/etudes/monstration/section4ok.shtml#grand2]

(3) Un exemple potentiellement important pour le musée concerne le fait que les navigateurs, modules externes et autres logiciels, même s'ils sont actuellement gratuits pour la plupart, ne sont pas libres de droit, car appartenant à des entreprises, et pourraient ne plus être mis à jour, devenir payants, etc. Si des standards de conservation se mettent en place, il serait souhaitable qu'ils soient basés sur des logiciels libres afin de les modifier directement, partager avec d'autres institutions, etc.

(4) « *The problem, of course, is that the features that would likely be lost in such a conversion would be exactly those features on which Web artists currently depend. Most art projects now being made for the Web rely not just on a specific collection of verbal or quantitative information, but on a carefully controlled visual design or play on the medium that enframes them. What will become of an artist's precisely rendered 800- x 600-pixel bitmap if screen resolution jumps to 10,000,000 x 10,000,000? It will degenerate into either a tiny image in the middle of an otherwise blank screen, or a big image with appallingly low resolution. This is not an academic question for museums that are considering acquiring Web projects into their collections.* » dans Ippolito, Jon. « The Museum of the Future : A Contradiction in Terms? » (en ligne), *Artbyte*, n° 2, juin-juillet 1998, p. 18-19. [http://three.org/ippolito/writing/wri_cross_museum.html].

(5) [http://www.potatoland.org/shredder/shredder.html]

(6) Variable Media en anglais. [http://www.variablemedia.net/]

(7) « *As outlandish as the idea may seem to traditional collecting practices, the Variable Media Initiative offers an alternative for those whose conception of their work goes beyond its manifestation in a particular form. And it helps us imagine the museum as an incubator for living, changing artworks — rather than a mausoleum for dead ones.* » dans Ippolito, Jon. *The Museum of the Future : A Contradiction in Terms* [http://three.org/ippolito/writing/wri_cross_museum.html]

(8) dans Boutteville, Joël, Annick Bureaud et Nathalie Lafforgue. op. cit. section IV.II.3.

Bibliographie

Ouvrages, rapports :

Bolter, Jay David et Richard Grusin. *Remediation - Understanding New Media*, MIT Press, Cambridge, Massachusetts, Londres, Angleterre, 2001, 4e édition, 295 p.

Bowker, Geoffrey C. et Susan Leigh Star. *Sorting Things Out: Classification and its consequences*, The MIT Press, Cambridge, Massachusetts, Londres, Angleterre, 1999, 377 p.

Bureaud, Annick et Nathalie Magnan (sous la direction de). *Connexions : art, réseaux, médias*, École Nationale Supérieure des Beaux-Arts, Collection Guide de l'étudiant en art, Paris, France, 2002, 642 p.

Boutteville, Jöel, Annick Bureaud et Nathalie Lafforgue. *Art et technologie : la monstration*, Délégation aux Arts Plastiques (DAP), Ministère de la Culture et de la Communication, Paris, France, 1996, environ 200 p. Aussi sur Internet à [http://www.olats.org/livresetudes/etudes/monstration/index.shtml]

Curating New Media, Third Baltic International Seminar 10-12 May 2001, Baltic/CRUMB, Angleterre, 2002, 198 p.

DEPOCAS, Alain, Jon IPPOLITO et Caitlin JONES (sous la direction de). *L'approche des médias variables : la permanence par le changement*, The Solomon R. Guggenheim Foundation, New York, États-Unis et la Fondation Daniel Langlois pour l'art, la science et la technologie, Montréal, Canada, 2003, 138 p. Aussi sur Internet à [http://switch.sjsu.edu/web/v6n1/article_a.htm]

KISSELEVA, Olga. *Cyberart, un essai sur l'art du dialogue*, L'Harmattan, collection Ouverture philosophique, Paris, France, 1998, 368 p.

Modern Art : Who Cares? An interdisciplinary research project and an international symposium on the conservation of modern and contemporary art, sous la direction de IJsbrand Hummelen et Dionne Sillé, The Foundation for the Conservation of Modern Art et le Netherlands Institute for Cultural Heritage, actes du International Symposium on the Conservation of Modern and Contemporary Art, 8 au 10 septembre 1997, Amsterdam, 1999, 445 p.

ROTHENBERG, Jeff. *Avoiding Technological Quicksand : Finding A Technical Foundation For Digital Preservation*, Report for the Council on Library and Information Resources,Washington, 1999, 39 p. Aussi sur Internet à [http://www.clir.org/pubs/reports/rothenberg/contents.html].

Articles sur Internet

BESSER, Howard. *Longevity of Electronic Art* [en ligne], présenté à International Cultural Heritage Informatics Meeting, 3 au 7 septembre 2001, Milan (Italie), 2001. [http://www.gseis.ucla.edu/~howard/Papers/elect-art-longevity.html].

BROCHU, Danièle et Jean DAVALLON. « L'organisation du musée confrontée à l'arrivée des nouvelles technologies », Médiations sociales, systèmes d'information et réseaux de communication, 11^{e} Congrès national des sciences de l'information, Metz, 1998.

DIETZ, Steve. « Memory_Archive_Database v 3.0, 2000 » [en ligne], *Switch*, n° 13, San Jose State University. [http://switch.sjsu.edu/web/v5n3/C-1.html].

IPPOLITO, Jon. « The Museum of the Future : A Contraction in Terms » [en ligne], *Artbyte*, n° 2, juin-juillet 1998, p. 18-19. [http://three.org/ippolito/writing/wri_cross_museum.html].

LIALINA, Olia. « Un lien serait déjà assez » [en ligne pour les abonnés], *Archée*, Montréal, juin 2001. [http://archee.qc.ca/ar.php?page=imp&no=163].

Manovich, Lev. *Metadata, Mon Amour?* [en ligne], 2002. [http://www.manovich.net/DOCS/metadata.doc].

RINEHART, Richard. « The Straw that Broke the Museum's Back? Collecting and Preserving Digital Media Art Works for the Next Century »[en ligne], *Switch*, n° 14, San Jose University, juin 2000. [http://switch.sjsu.edu/web/v6n1/article_a.htm]

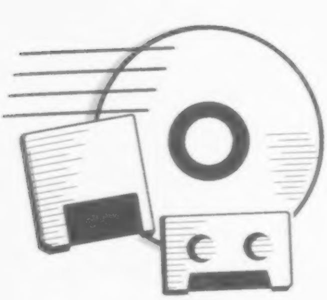
RINEHART, Richard. « Preserving the Rhizome ArtBase » [en ligne], *Rhizome*, septembre 2002. [http://rhizome.org/artbase/report.htm].

ROSS, David. « Net.art in the Age of Digital Reproduction » (en ligne, transcription d'une conférence donnée à la San Jose State University, États-Unis, 2 mars 1999, *Switch*, 1999. Sur Internet [http://switch.sjsu.edu/web/ross.html].

Abstract

Net Art developed with the advent of the Internet, and since that time has been drawing the attention of museums. The museums that display, order, and purchase Net Art are now faced with developing a preservation strategy specific to it. Even though these museum may have extensive experience in the conservation of contemporary art (which leans more and more, or at least in part, towards ephemeral art), Net Art presents unique challenges.

Net Art works are, by definition, ever-changing and depend on numerous parameters to be seen. Various approaches to their preservation are currently being explored, including those used by the Variable Media Network of the Guggenheim Museum in New York and the Daniel Langlois Foundation in Montreal, where artwork is no longer defined by its medium and is viewed as a work in progress rather than as a finished product. For now, the best approach to preservation is probably a combination of the museum strategy with automatic and continuous Internet archiving.



RECORD, PLAY, FAST FORWARD: DEVELOPING STRATEGIES FOR THE CARE OF ELECTRONIC MEDIA ART AT THE ART GALLERY OF WESTERN AUSTRALIA

VANESSA GRIFFITHS

Art Gallery of Western Australia
PO Box 8363, Perth Business Centre
Perth WA 6849
Australia
Tel.: 08 9492 6733
E-mail: v.griffiths@iinet.net.au

Abstract

Artists have increasingly made use of the vast possibilities presented to them by multimedia and digital technology, and will continue to do so. But the prospect of how to preserve these artworks — how to archive the material, how to maintain the equipment, how to document an installation, and how to overcome the inevitable obsolescence of the technology — is a glaring problem for art galleries. This situation has been recognized by conservation, curatorial, and registration staff alike and practical solutions are now being sought to improve the physical care and documentation of electronic media technologies.

Introduction

> Marcell Duchamp has declared readymade objects as art, and the Futurists declared noises as art — it is an important characteristic of my efforts and those of my colleagues to declare as art the total event, comprising noise/object/movement/ & psychology — a merging of elements, so that life (man) can be art.
>
> — M. Vorstell, p. 177 in Rush (1999)

The motivation for this paper stems from an interest in and enjoyment of contemporary art and its usage of new technologies. The way in which the entire mix can meld with conservation is challenging and provides conservators with a broader technical community in which to engage.

Sometimes the pace of technological change far outstrips our capability to comprehend and develop skills to maintain what artists create with contemporary technologies. Artists have increasingly made use of the vast possibilities presented to them by multimedia and digital technology, and will continue to do so. At the Art Gallery of Western Australia (AGWA), which is a state institution, we have been collecting artworks that include videotape, sound recordings, electronic media, and more recent technologies such DVD for more than a decade. Even though our conservation staff, registrars, curators, and collection managers are experienced with the conservation issues that arise from more traditional material, the area of electronic media and new technology artworks tends to fall into a type of conservation "no-mans land." For example, where do such types of artworks fall within conventional conservation disciplines? Is it practical to dissect these artwork into familiar components for condition reporting? And how in fact does one condition-report video-based installation art?

One of the most glaring problems for art galleries, in general, is the prospect of how to preserve these artworks — how to archive the material, how to maintain the equipment, how to document an installation, and how to overcome the inevitable obsolescence of the technology. This situation has been recognized by conservation, curatorial, and registration staff alike and practical solutions are now being sought to improve the physical care and documentation of electronic media technologies.

"Record" — Understanding electronic-media-based installations

Description of the collection

AGWA has a relatively small number of artworks that contain analog and digital formats, including VHS, film, U-Matic, Beta SP, and, more recently, DVD and laser disc. The majority of these artworks have been classified by the curatorial discipline as contemporary sculpture or installations, primarily because the video or filmic image is only one element of the varied materials that make up the entire artwork. Although the number of these installation-based artworks is small, they are frequently on display because of their uniqueness and entertainment value.

Typical of the type of artwork collected are works by artists Julie Rrapp and Sooja Kim where the image(s) are projected onto a wall or screen or played on a monitor. For example, Korean artist Sooja Kim's *Sewing into walking* features a three-channel video installation incorporating one video projector, three monitors, and bundles of used clothing wrapped in bed sheets placed on the floor in front of the projected image; it has been exhibited three times since its acquisition in 1998. Another example, acquired in 2001, is Rodney Glick's *Earthquake* which features two images that are recorded onto DVD and projected onto a wall simultaneously; this is accompanied by a soundtrack that has to be amplified to a level that requires the exhibition space to be almost completely soundproofed. Both of these works promote contemporary art as a form of entertainment, and provide an almost interactive experience for the audience. Other examples are Julie Rrapp's *Rise and fall* and Carol Rudyard's *Salt cellar and glass*, both of which include associated materials such as seats, tables, artificial plants, floor tiles and moving ballet shoes, and a rotating television set.

The practical gap — New technologies, new conservation problems

There seems to be a consensus in the conservation community that electronic media artworks can suffer badly from "under documentation." And if a technology-based artwork is condition-reported in the same manner as a bronze sculpture or oil-on-canvas painting, the technology-based installation documentation will definitely be lacking in detail. Unfortunately, the problems presented by the technology-based installation are often far outside the experience or the expertise of the conservators responsible for their care (Bishop 2001). Coupled with a need for formal training programs that specifically focus on contemporary art, conservators responsible for this type of material will also need the ability to extend the traditional static methods of aesthetic documentation to accommodate multimedia technology and the ability to develop new methodologies for undertaking the documentation. As Laurenson (1999a) points out in her comprehensive discussion of the conservation of Gary Hill's installation *Between cinema and a hard place*, conservators will need to include in their vocabularies the idea of "mechanical breakdown and obsolescence of parts or whole technologies." Conservators will need to be able to recognize what elements of an artwork are artistically essential to the meaning of the artwork, what about it can be changed, and what cannot be changed. To illustrate this point, consider the stack of three monitors in Sooja Kim's *Sewing into walking*. These monitors are a sculptural element of the artwork and should not b considered any less. However, if the condition report describes only the condition of the associated bundl of fabric that are placed nearby on the floor and ignores the other elements, the artwork is only partially documented.

Contemporary vulnerability

Artworks containing electronic media can be just as fragile and intrinsically unstable as other unique objects and paintings. For example, VHS tapes can become damaged and the quality of the images can deteriorate if the storage conditions have been poor; and this is applicable to most analog formats. Numerous copies may have been made from the acquisition format on an ad hoc basis, complicating both conservation and registration. Components suc as playback equipment and hardware may deteriora and/or suffer mechanical failure over time, and parts and technical expertise may become scarce as manufacturers of technology cease to produce critica components. In short, although artworks containing electronic media are different from other artworks (as they need to be installed for exhibition and rely on the support of someone knowing how all the components fit together), they can be equally susceptible to damage.

It is also important that conservators be aware that "the life of an electronic artwork can be understood to be a fluid one that evolves as video formats chang and components deteriorate or become obsolete due to age or market fluctuations" (Sterrett and Graham 1997). But conservators are indoctrinated through their training to prevent change to artworks, and the thought of changing some part of an artwork can cause initial angst. The best response to this "fluidity is to record and document as many aspects of the artwork as necessary to maintain the "essence" or the artistic intention of an artwork regardless of wha format the technological component takes; this will allow the artwork to be reconstructed faithfully each time it is shown (Sterrett and Graham 1997).

Ethics — Going with the flow

To overcome obsolescence it must be recognized that flexibility and fluidity are part of contemporary art. For example, there have been instances in the past at AGWA where technology formats in various installations have been transferred or migrated to different formats. This is not necessarily a drastic event, but the intervention and reason for it need to be documented just as any change for other types of artworks would automatically be recorded. For example, when undertaking a previous survey of the technology-based artworks and the number of formats held in our state collection, it became apparent that there had been random, undocumented interventions on a number of artworks. This creates a very frustrating situation for conservators. Without being able to establish if these changes were done in accordance with the artist's wishes or were simply the whim of a technician who decided that DVD would provide a cleaner image than VHS tape, a framework for understanding the full condition of the technology used for that particular artwork cannot be gained.

Also found during the survey was that few of the items were identified as being a "master copy" or an "exhibition copy," nor were dates of production given. This has a direct impact on registration and conservation disciplines, and again relates to the under-documentation of technology-based artworks if there is no formalization of capturing this process. This situation is not unusual in art galleries and I have heard of similar instances in other collections from conservators who face the same problem. One explanation for this is given by Mark Roosa (1998):

> The reason for this is that the medium is relatively young compared to traditional art forms. Another reason why documentation is scarce is because much of what is known about the maintenance of installation hardware resides outside of the conservation communities but with electrical engineers, audio and video specialists and technicians.

The motivations for these actions are usually practical and mostly driven by the demands of exhibitions to have reliable or numerous copies of the necessary format on hand. But are these interventions always ethical? To draw a parallel with paintings conservation, would it ever be appropriate to repaint or reproduce a painting because the quality of paint has been improved?

One ideology is that if the actual playback equipment or hardware cannot be seen or does not affect the intrinsic artistic intention of the artwork, then there should not be a problem with the migration from one format to another, e.g. from audiotape to digital mini disc. An example of this particular issue in AGWA's collection would be Ronald Jones' *Untitled: Edward Earl Johnson sat in this chair as he ate his last meal*. This sculpture is fabricated from a range of materials, and comprises a single aluminium chair standing on a tall plinth while a "music box" style song plays softly. The song is *Always* by Atlantic Star, and it is a highly significant element of the sculpture (this song was sung by Edward Earl Johnson's family outside the prison just hours before his execution in Mississippi State Penitentiary in 1987). All of the playback equipment and speakers are concealed within the interior of the sculpture; the speakers are in the seat of the chair and the actual playback equipment is hidden inside the plinth. The original format of the sound recording was magnetic audiotape played on a cassette player, but it became evident after some months on display that the sound quality was deteriorating and mechanical breakdowns were becoming more frequent. The decision was therefore made to migrate the sound recording onto a digital format. The delivery system now comprises a commercially available mini disc player and mini discs. The mini disc player is not visible and does not affect the exterior appearance of the artwork; it does, however, allow the audio element of work to be delivered by a much more reliable technology.

"Play" — Playing installation-based artworks

Towards new documentation methodologies

How is it that stage plays and musicals can be re-created or screenplays re-made? The simple fact is that somewhere, someone noted or accumulated the information necessary to allow a director to re-create it in some form in the future. The strength of the future performance and how accurately it matches the original depends on the supporting documentation. A parallel logic can be applied to the documentation of contemporary technology-based installations, which face the risk of not being installed properly or authentically in the future unless they are properly documented (Laurenson 1999b):

> They depend on those responsible to manage a complex array of detail and to know how the work is to be shown. Until recently, most of the conservation literature regarding video and electronic media, was aimed and structured specifically for library and archive collections which contain large volumes of video, film and sound recording etc.

This statement accurately describes the fact that library and archival collections are very different than art museum/gallery style collections, and require different methodologies and ideologies. One major difference between the two is that libraries are often equipped to deal with the storage of large amounts of information, the majority of which needs to be readily accessible in a number of formats, whereas art collections focus on the uniqueness of objects and have a tendency to believe that there can only ever be one physical manifestation to make the artwork authentic. The level of documentation necessary to accurately re-create technology-based installations rapidly becomes apparent through day-to-day workload and exhibition scheduling, where it is often the collective memory of staff rather than written documentation or even the artist's input that guides the installation. This collective memory skill might be admirable but it does not ensure the correctness of future installations.

In an effort to improve the level of documentation carried out, conservation and registration staff at AGWA have put together two procedures to improve on the acquisition and documentation processes. The first is a basic questionnaire for artists when a technology-based artwork is acquired, with questions arranged into the following categories:

- Original format
- Equipment
- Display/installation
- Copyright and photography
- Conservation

Although the questionnaire provides some information regarding the works, it is not all-encompassing nor can it be relevant if the artist is no longer alive to provide the information. The second procedure is a condition-report template specifically for technology-based artworks; this template includes the establishment of what physical components are part of the artwork and a list of everything associated with the work.

Improvements in the acquisition process

The acquisition process at AGWA has also begun to change in that when the gallery acquires an artwork with a technology element, it now requests that the artist provide four copies of the format on which the artwork is contained (although there is no attempt to dictate to the artist which technology to use). The four copies are used as a master copy, a sub-master copy from which other copies can be made, an exhibition copy, and a loan copy, and it will be AGWA's responsibility to decide how best to archive the material. Using this registration process, the artist and the curator proposing the acquisition can now establish whether or not the artwork is an edition and if there is an artist's proof, which may or may not be sold.

Understanding the meaning of documentation

The documentation of contemporary art that involves new technologies is not just about establishing a long list of facts about the artwork (Kraemer 1999). The questions and answers must be designed so as to allow the development of an understanding of the meanings and functions of the work of art. The artist's own thoughts on the deterioration or obsolescence of some aspect of the artwork should also be sought. The condition-report format that AGWA has developed for technology-based art follows closely the guidelines and templates that the Tate Modern Art Gallery has established for the documentation of its electronic media collection. These practical actions are hugely important in working towards developing a condition-reporting methodology. The information that conservators are trying to collate include floor plans (written either by AGWA staff or supplied by the artist), installation instructions as supplied by the artist, diagrams, lists of necessary equipment, suppliers, and manufacturers, and installation photographs.

The installation of Derek Kreckler's *White Pointer – You are listening to the sounds of humans observing fish*, which features various components and audio equipment, provided a typical example of re-creating an artwork in a format that was different than the original version (the artist made a deliberate choice to update from magnetic audiotape to digital recording on mini disc). The installation of this artwork at AGWA provided an excellent opportunity to discuss aspects of the work with the author and to apply the new methodology of documenting installation art. In this case the artist uses sound to delineate space, so the distance between the stations and the volume setting for the five different soundtracks is very specific; and because this information is so importan to the artwork, it is vital that it be captured in the documentation if the artwork is to be re-created accurately in the future. Through the documentation process parameters, it is possible to establish which aspects of an artwork are flexible and which are not. As artworks are sometimes re-created in different venues, harsh realities such as room size, ceiling height, light levels, and exhibition time frames can all encroach on the re-creation of the work. For example, Sooja Kim has stated that her work *Sewing into walking* can have a certain amount of flexibility i terms of size of the room and projection space so tha it can be integrated into the ongoing display program at AGWA.

"Fast forward"

The next step

So, fast forward to the future. Will AGWA be able to install or "play" its technology-based installation artworks in 10 years time? Although it is clear that improvements in the areas of documentation and acquisition have begun, there is still the obvious need for strategies and action, particularly on the issues of archiving the material that is already held and risk analysis. And, of course, because most technology is driven by consumerism and manufacturing, there is no way to tell what technologies will become obsolete or what form of technology artists might employ next.

Widening of professional resources

The capability to develop broad communication between relevant professionals in the information technology, computer science, and film and video industries will become increasingly important in the preservation of technology-based artwork. AGWA already has good working arrangements with other local galleries and institutions when extra equipment for installations is needed for an exhibition. But installations often also require electricians to install complex wiring, experts in the film and video industry to provide advice, and commercial operators to make copies of formats when needed. As art galleries do not usually have the staff and equipment to transfer formats, archive material, make copies, create digital formats, or even maintain equipment, it is likely that outsourcing these activities to commercial service providers will become commonplace.

Conclusion

The profession of conservation is currently faced with the challenge of preserving and continuing to exhibit a variety of technology-based artworks, although whether or not this will become a specialization of its own remains to be seen. At AGWA, the staff are working towards improving the documentation and understanding of new technology artworks and installations by gaining more information from the artist and curator and establishing routines and consistency. These small improvements are encouraging and, although they won't solve the problem overnight, they will ensure that professional ethics and standards are less likely to be compromised. As art pushes into new realms of media, collecting institutions must to be prepared to examine the cost to care for these artworks, and develop the necessary methodologies, professional networks, and skill bases. By staying open to the conservation challenges of new technologies and media art, the long-term usability of the artworks may well be ensured — or at least we can avoid having to press the "Delete" button!

Bibliography

Bishop, M.H. "Evolving Exemplary Pluralism: Steve McQueens's *Deadpan* and Eija-Liisa Ahtila's *Anne, Aki and God* — Two Case Studies for Conserving Technology-based Installation Art. *Journal of the American Institute for Conservation* 40, 3 (1999), pp. 170–191.

Digital Preservation Coalition. www.dpconline.org

Graham, J., and J. Sterrett. *An Institutional Approach to the Collections Care of Electronic Art*. 1997. aic.stanford.edu/conspec/emg/emsig_san_diego.html (accessed 08/09/2002).

Jones, M., and N. Beagrie. *Preservation Management of Digital Materials: A Handbook*. London, UK: The British Library, 2001.

Kraemer, H. (University of Cologne, Media and Cultural Communications, Cologne D 50969, Germany). Personal communication, 1999.

Laurenson, P. "Developing Strategies for the Conservation of Installations incorporating Time Based Media with reference to Gary Hill's *Between cinema and a hard place*." *Journal of the American Institute for Conservation* 40, 3 (1999a), pp. 259–266.

Laurenson, P. "The Conservation and Documentation of Video Art." pp. 263–271 in *Modern Art: Who Cares? An Interdisciplinary Research Project and International Symposium on the Conservation of Modern and Contemporary Art*, edition 1 (edited by I. Hummelen and D. Sille). Amsterdam, Netherlands: The Foundation for the Conservation of Modern Art and the Netherlands Institute for Cultural Heritage, 1999b, 445 pp.

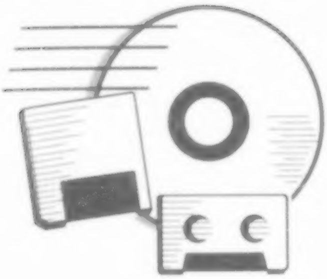
Real, W.A. "Towards Guidelines for Practice in the Preservation and Documentation of Technology-based Installation Art." *Journal of the American Institute for Conservation* 40, 3 (1999), pp. 211–231.

Roosa, M. "Maintaining Technology Based Installation Art." pp. 30–47 in *Playback: A Preservation Primer for Video* (edited by S.J. Fifer, T. Gould, L. Hones, D.H. Norris, P. Ramey, and K. Weiner). San Francisco, CA: Bay Area Video Coalition, 1998.

Rush, M. *New Media in Late 20th-century Art*. London, UK: Thames and Hudson, 1999.

Résumé

De plus en plus, les artistes font appel aux énormes possibilités que leur offrent le multimédia et la technologie numérique, et ils continueront de le faire. Mais l'un des problèmes les plus flagrants auxquels se heurtent les musées est le mode de conservation de ces œuvres : comment archiver le matériel, conserver l'équipement, rassembler des données sur l'installation et surmonter l'inévitable difficulté de la désuétude de la technologie. Les restaurateurs, les conservateurs et les registraires, conscients de cette situation, sont à la recherche de solutions concrètes pour améliorer l'entretien et la documentation des œuvres sur support électronique.



PRESERVING RESEARCH DATA: A TIME FOR ACTION[1]

CHARLES HUMPHREY

University of Alberta
Edmonton AB
Canada
E-mail: Chuck.Humphrey@ualberta.ca

Abstract

Recent publications about the preservation of electronic records have all but ignored "research data" as a class of digital objects warranting attention. One explanation for this neglect is general confusion about what research data are. The work of the National Data Archive Consultation (NDAC) in Canada provides a definition of research data that distinguishes these records from all other digital objects. This paper discusses the importance of this definition and reports some of the findings from the NDAC, which has produced the latest knowledge about the current risk level of research data in Canada. A gap analysis showed that research data fall outside the current interpretation and execution of the mandates of Canadian heritage institutions. Furthermore, evidence reveals that data loss becomes acute without an institution with the specific mandate to preserve research data. Finally, without a recognized institution responsible for data preservation, researchers do not know where or how to archive the data from their research.

Introduction

A proliferation of position papers on digital preservation appeared in the mid-to-late 1990s in conjunction with the growth of the digital library movement. Quite bluntly, digital preservation has not been a neglected topic. For example, the December 2001 background report of the Canadian Association of Research Libraries (CARL/ABRC) on the preservation of digital resources lists more than 40 documents published between 1996 and 2001.[2] This selected bibliography is international in scope, including titles from Australia, Canada, France, the Netherlands, the United Kingdom, and the United States. A variety of institutions sponsored these works, including the National Library of Australia, the National Library of Canada (NLC),[3] Bibliothèque National de France, Koninklijke Bibliotheek, the Library of Congress, and the Library and Information Commission. Several organizations and associations contributed reports, including the Council on Information and Library Resources (CILR), the Research Libraries Group, the International Council for Scientific and Technical Information, InterPARES, the IBM Almaden Research Center, and the Online Computer Library Center (OCLC). Yet within this generous body of literature about digital preservation, only one title in the CARL/ABRC bibliography addresses research data.[4] Seemingly, the presence of research data in the dialogue on digital preservation has been almost invisible.

A natural outcome of this rich international interest in digital preservation was a movement within UNESCO in 2001 to establish a charter on the preservation of digital heritage.[5] In September 2002, consultations were initiated with a wide range of stakeholders to gather critical comments on the preliminary draft of this charter. In a joint response, the International Association of Social Science Information Service and Technology (IASSIST) and the International Federation of Data Organizations (IFDO) addressed a conspicuous omission in the initial version. Specifically, research data were not mentioned in any of the proposed charter's declarations. The joint submission from these two organizations offered several suggestions on how to incorporate research data into the charter. At least one of these ideas was adopted because the version released following the consultation now includes a clause calling upon member states to designate one or more agencies to undertake the coordinating responsibility for the preservation of research data.[6]

Overall, the literature on digital preservation rarely makes reference to research data, instead focusing its spotlight on digitally published materials. An example of this perspective is found in the report of the United States' National Digital Information Infrastructure and Preservation Program.[7] Six categories of digital content are explicitly named: e-books, e-journals, digital music, digital television, digital video, and Web sites. With the exception of Web sites, the other five categories represent digital versions of traditional content. Where does this leave research data? Are research data really that different from other digital content? Is it because people are unsure about what research data are? How does one identify research data?

The discussion below shows the relationship between research data and digital heritage. It further provides a definition distinguishing research data from other digital content. While the issues associated with the preservation of research data are similar to other digital content, research data are different in significant ways. These differences unfortunately seem to isolate research data and, unfortunately, to place it at greater risk. Evidence is presented about the tenuous situation of research data in Canada and an assessment is provided about its level of vulnerability. Information gathered through the NDAC[8] is included in this discussion as it relates to these topics.

Springing forth from everything that is digital?

How do we distinguish data from non-data in the digital realm? Little insight is gained from popular usages of "data," even though one might expect that, living in the digital age, a precise meaning of this concept would exist. From the perspective of the average person, data consist of anything that supports or underlies information. The meaning is further blurred because information and data are often used interchangeably. Information is data; data are information. Consequently, clearly communicating the objects of data preservation is complicated by a lack of clarity about what, precisely, is meant by the term "data."

In addition to many popular usages of "data," a variety of technical definitions exist. In the Webopedia, several definitions relate data to various aspects of computing.[9] According to one definition, data are distinct pieces of information formatted in special ways for manipulation by programs. A narrower version of this definition applies to programming, where data are the binary values that represent symbols and machine instructions. From this outlook, to be digital is to be data, which is an unpromising start to delineate data objects for preservation.

Treating everything that is digital as data quickly proves futile. Simply too much of our world is digital for this to be practical. The authors of *How Much Information?* point to the dominance of today's digital content.[10] The combination of a significant decline in the cost of massive digital storage and of the rapid growth in the volume of information that is "born digital" has made digital information the dominant form of information production. Lyman and Varian estimate that between one and two exabytes (which is a billion gigabytes) of unique information is produced each year. They estimate this to be 250 megabytes for every man, woman, and child on earth.[11]

When the NDAC Working Group began their study into the need for data archiving services in Canada, a variety of ideas quickly surfaced about definitions of "data." The Working Group consisted of experts from the humanities and the social sciences,[12] which contributed to differences in understanding. Are digital images — still or motion — data? Is digital audio data? Are electronic texts data? Is e-mail data? Is all statistical information data? Differences also existed within disciplines. Some viewed a copy of a book in electronic format as data. For others, a copy of this work had to be encoded in a mark-up language before it qualified as data. For some, an electronic table from a survey was considered data, while others recognized only the raw file from which the table was originally constructed as data. If format alone is insufficient to define data, what criteria are useful?

The NDAC Working Group chose three general characteristics to define research data, which are described in the following statement (p. 4):

> [A] research data archiving function is defined as preserving, managing and making publicly accessible digital information structured through methodology for the purpose of producing new knowledge. Such an archiving function provides stewardship for those outputs of the research process that exist between raw research materials and published results. Acquisitions could include digital information produced by researchers and of interest to researchers.[13]

First, methodology plays a key role in distinguishing research data. This attribute helps separate raw data from research data. Raw data are for the most part unorganized and unrelated pieces of information, while research data have a meaningful structure imposed by a methodology. This principle is readily recognized in science, where replicable methodologies are used to govern the creation of data. People, for example, have opinions. Through survey methodologies, these opinions can be gathered and organized into research data. A text in simple electronic format is raw data. However, applying a mark-up methodology and incorporating tags in the text will convert it into research data. The outcome of organizing raw information according to a defined method is the entity of interest for data preservation.

Another characteristic of research data is the potential that they have for the production of new knowledge. Evaluating the knowledge-creation utility of data may differ from assessments about their value to national memory. Instead of examining records only according to their importance as legal, administrative, or historical evidence, research data are also evaluated according to the contribution that the data may make to future knowledge. Research data have analysis value beyond their initial use. They are also important in fueling new discoveries through secondary analyses and through the creation of new data by combining existing data.[14]

A third attribute is associated with the close relationship between research data and the research community. The data of preservation interest are both those that have been compiled by researchers and those that are of value to researchers. The interests of researchers may well include digital objects that would not be identified as research data according to the previous two characteristics. For example, e-mail would be excluded as research data applying just the methodology criterion mentioned above. However, a researcher may utilize a collection of e-mail messages from a variety of sources to develop an algorithm to identify authorship. The interest in the content and the subsequent structuring of raw data through methodology may result in the creation of research data from other digital objects. This criterion also supports classifying some administrative databases, such as hospital records, as research data. Most administrative databases are not created with researchers in mind. Nevertheless, the contents of some administrative databases may be of great interest to researchers.

To summarize, research data were defined by the NDAC Working Group as digital objects shaped by methodology for the purpose of creating new knowledge. Research data consist of both data produced by researchers and data that are of interest to researchers. As part of the research cycle, they fall between the raw material of research and the published findings. Currently, institutions capture published outcomes of research, but no national Canadian institution has assumed responsibility to preserve research data.

Does this matter?

Once the Working Group had an operational definition of research data, an analysis of the institutional support for archiving such data was undertaken.[15] Looking at Canadian institutions with responsibilities for the preservation of heritage, the Working Group examined existing mandates and practices to determine the types of digital objects that are currently protected under the national preservation system and those that fall between the cracks.

The mandate of the NLC is to preserve the published heritage of Canada, which in recent years has been extended to electronic publications, electronic journals, and Web sites. Research data, however, are not considered publications and, consequently, are outside the current preservation net. The National Archives of Canada (NAC) has responsibility for preserving the records of the federal government and private records of national significance. While the NAC's mandate is general enough to encompass research data, in practice it takes in only a small sample of the federal and public records that exist. The Working Group determined that a wider and more representative sample of research data is needed than the NAC has the direction or resources to archive. In addition, some Canadian research data are international in scope, exceeding the boundaries of Canada. In these instances, the data may be outside the NAC's jurisdiction. The conclusion of this analysis was that the vast majority of academic and non-academic research data fall outside the current interpretation and execution of the mandates of the NLC and NAC.

The Working Group's investigation further showed that no other Canadian institutions, including university libraries or archives in other sectors, have a national mandate or the resources to address the current level of need for preserving research data (p. 11):

> Archivists from many sectors — federal, provincial, post-secondary, and public — testify that they cannot separately or together solve the massive, historic problem of data management without an overarching national data archiving function or agency. Many archives handle research data now, but none considers itself capable of providing the essential central infrastructure.[15]

This analysis substantiated the concern that no Canadian institution currently operates under a mandate to preserve research data. Knowing this, however, does not answer two related questions with significant political implications. Should we care that research data are not preserved? Is the problem large enough to justify doing anything?

Questions about whether or not it matters to preserve research data have been asked in other countries.[16] The answers found internationally apply equally to Canada. Some data require preservation as a condition of funding. The production costs of data tend to be very expensive.[17] To receive a greater return on this investment, funding agencies often require research data to be preserved for secondary use and analysis. Furthermore, when the costs of data are paid through the public purse, an additional obligation exists to ensure the proper protection of the public's investment.

There are also stewardship and custodial reasons for preserving research data. In addition to practicing good stewardship of public funds, researchers have an obligation to preserve the evidence upon which their research findings are determined. Just as research outcomes are preserved through a system supporting the preservation of publications, the data on which these findings are based need protection. Data collections often become the focus of specific developments within fields of study and end up extensively referenced in a discipline's literature. A custodial responsibility exists to protect the data upon which this literature is based.

Furthermore, ethical reasons exist to support the preservation of research data. The collection of data from human subjects by academic researchers requires the prior approval of research ethics boards. The primary focus of ethical considerations tends to be the privacy and protection of human subjects, but ethics boards must also weigh the burden of repeated data collection on rare or special populations. As a safeguard against over-interviewing and burdening these populations, data that are collected should be conditionally shared among researchers working with special populations. To be shared, data need to be preserved. A current debate within health research is about the amount of time that data should be retained before they are destroyed. One extreme approach taken by ethics boards in health research has been to stipulate a shelf-date after which data must be destroyed. This tactic to protect human subjects in health research is now being questioned. Preservation practices are being considered that allow saving research data without jeopardizing the privacy or protection of the subjects in this research.

Similarly, legal grounds exist to preserve research data. Legislation, such as the Statistics Act in Canada, emphasizes the protection of individuals and guards against the misuse of individual-level data that might lead to the disclosure of a person's identity. Strict conditions of access to confidential data, however, do not obviate the need to preserve this information through proper archival practices. The argument can be made that the process of preserving confidential data can increase its level of security. Preservation can validate the conditions under which access to data are allowed.

A primary reason for preserving research data, already mentioned above, is the potential that data hold for the creation of new knowledge. The value of data has already been stated both in terms of the monetary investment in its creation and in the amount of knowledge transfer based upon it. The multiple ways of creating new data from old, however, bring additional value and importance to the preservation of research data. The construction of time-series from earlier collections of data files is a fairly well-known example of this kind of activity. The possibilities also exist of generating new units of analysis by combining data sources for which a common unit exists, but around which the data were not originally collected. For example, a variety of data files may have been created at the level of individuals receiving medical care. Included in these files may be data on the hospitals where care was received. The original unit of analysis was individual patients. However, a new research project may wish to use the institution or the hospital as the unit of analysis. The creation of new data combining information across files for a new unit of analysis is only possible if the old data have been preserved.

The reasons given to justify preserving research data are not just for a few small areas of research, either. This is a pervasive problem extending across disciplines and sectors in society. In the Phase I Report (p. 14) of the NDAC, a wide range of stakeholders identified the need to protect against the loss of data across entire classes of research:

- geo-spatial data
- social science and humanities data sets created with Social Sciences and Humanities Research Council of Canada (SSHRC) funding
- social science and humanities data sets funded from other sources
- unpublished data created by the federal government
- unpublished data created by provincial, municipal, or private-sector sources
- statistical survey data
- government or private polling data

The above arguments present a strong case for preserving research data on the basis of the financial investment, the custodial responsibility, the legal and ethical obligations, and knowledge generation.

In addition, a comparison of Canada to the countries used by our leaders to rank Canada's performance reveals that most of these nations already have institutions responsible for archiving research data.[18] This poses a question about Canada's competitiveness with these nations in the international knowledge economy. For example, one study suggests that Canada's participation in the Global Change program (an international science program built around the exchange of scientific data) in the 1990s was weakened because no institution exists in Canada to preserve research data.[19]

Evidence of the severity of the problem

How significant is this problem in Canada? Evidence from two Canadian studies provides insight into the seriousness of not having national data archiving services. First, a survey conducted in conjunction with the NDAC shows that a substantial number of the research projects funded by the SSHRC between 1998 and 2000 involved the production of research data.[20] Furthermore, the vast majority of these researchers had no experience archiving data nor did they have plans to preserve the data from these projects. The second study shows that more than 97% of the research data produced by SSHRC-funded projects between 1978 and 1981 cannot be located today. The data from the three projects that were found are all held by the Archive of the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan. The data from the remaining projects may be in the hands of individual researchers but the chances that these data are recoverable are very low. Details about these two studies are provided below.

How large is the problem?

A 20%, stratified sample of all recipients receiving a standard research grant from the SSHRC between 1998 and 2000 were interviewed about their experiences in creating and archiving research data and about their use of data created by others.[21] Respondents were also asked a set of questions tapping attitudes toward the archiving and sharing of research data. A 26% response rate (116 respondents) was achieved.

The results of this survey show that a substantial percentage of the projects conducted during this 3-year period involved the creation of data products. Fifty-five percent of the respondents said that they had created data files or a database as part of their most recently funded project. Of this group, only 7% said that they had deposited data from the project with an archive. However, a subsequent question revealed that of this 7%, most of them had not deposited data with an actual archive. Their answers showed confusion about what qualifies as an archive.[22] For example, one researcher felt that placing the project's data on the Web constituted an archival deposit. Respondents were also asked about the creation of research data in previously funded projects. In this situation, 59% reported having created data. Clearly, more than half of all researchers were engaged in the creation of data, either in their current research or in an earlier project.

Using these results, a conservative estimate of the confidence interval shows a rate of between 505 and 595 out of every 1000 standard grants involves the creation of data or databases as part of the project. Over the 3-year funding period of this study, an estimate of between 1100 and 1300 projects produced research data.

This study did not assess the nature of the data created. Therefore, no information exists about the size or scope of these data collections. The second study, which is discussed below, suggests however that the data associated with these projects are one of three types:

- original data collections that cannot be recovered if lost
- data created from other sources that exist as permanent records; these data could be reproduced if lost, although at the additional expense of re-creating the data
- data that are used from other providers, who have the custodial responsibility for the data

It is likely that the data from this 3-year period fall into these three categories. Consequently, the actual number of projects with data to preserve would be less than the interval between 1100 and 1300. Nevertheless, the results point to a sizable body of data that currently is not being captured for preservation.

This small survey does demonstrate the larger scale of the problem, however. These estimates are for just one research-funding agency and appear to be only the tip of the iceberg. The problem will only increase, likely exponentially, with the inclusion of projects outside the SSHRC standard grant stream. For example, no one has determined the amount of data generated through the research conducted in the Metropolis project[23] or other granting agencies, such as the Canadian Institutes of Health Research. Furthermore, no accurate estimate exists of the amount of research data produced by federal departments and agencies or their provincial counterparts.

How large is the risk?

While preparing a submission to the 1998 Consultations on the Future Role of the NAC and the NLC, an administrative study was uncovered that had been conducted in 1981 by the now-defunct Machine Readable Archives (MRA) Division of the Public Archives. This study was undertaken to locate SSHRC-funded research projects in which machine-readable data products had been created. From this list of projects, the dollar value of this research was estimated. The study consisted of a review of all SSHRC grants of CAN$5000 or more from 1977 to 1980, which happens to encompass the first 3 years of operation of the SSHRC. In a memorandum[24] to the Director of the MRA Division, John Saikkonen reported that, using their selection criteria, 150 projects were identified.[25] The total public expenditure on these grants exceeded CAN$7 million. In the conclusion of his report, Mr. Saikkonen observed, "it can be seen that the SSHRC is granting substantial sums for projects of national importance which have machine readable components."

Although the MRA Division was disbanded in 1986, the data coding sheets used in this study were preserved on paper allowing the reconstruction of the original data file. Included among the information gathered about each grant was its SSHRC file number. This set of SSHRC files, which represents a pre-identified population of funded projects in which there were computing expenditures, became a population of projects worthy of being examined 20 years later for evidence of data files and databases. A search for data from these projects was conducted in 1998 to see if a copy of these files could be found.

Of the initial set of 150 studies, 110 involved social science research while the remaining 40 were from projects in the humanities.[26] Using the names of the principal investigators and the titles from the 110 studies in the social sciences, a search of several data catalogues was conducted to look for evidence that data from these projects had been archived and were still accessible. Data from only three studies were found and all three of these were housed with the ICPSR in the United States.[27]

In November 2000, approval was given to investigate the original SSHRC files for a sample of 63 of the original 150 projects to see if new information could be learned about the data in each project. As grant-funded projects are finished, final reports and copies of publications from the research are added to each project's file. Many of the studies examined by the MRA investigators would not have been completed at the time of their investigation. New information collected from these files indicated that 3 of the 63 had no relevant data associated with the research even though computing expenses had been itemized in the project's budget, which affirms the effectiveness of the original selection criterion. A second finding consists of the three categories of data sources mentioned above, namely, original data, data produced from permanent records, and data obtained from other data producers. No further evidence was discovered in these files about intentions or plans to preserve data from these projects.

The findings from the follow-up investigation to the 1981 MRA administrative study reveal the grim results feared by the community concerned about preserving research data in Canada. Research data deposited with an institution responsible for preservation are still available 20 years later. In this case, the data from the three projects deposited with the ICPSR are still accessible. While the final status on the data belonging to the remaining studies is unknown without contacting the principal investigators, the chances that the data have survived are very slight. In other words, the risk of data loss is very high without an institution with the specific mandate to preserve research data.

Conclusion

The research conducted by the NDAC provides the most current and comprehensive evidence about the importance of a national institution with a mandate and means to preserve research data. A submission from the Canadian Association of Public Data Users in 1998 to the Consultations on the Future Role of the NAC and the NLC carried a message that resonates with these findings:

> The value of Canadian research data lies in its scientific worth, in the public investment expended on its creation, and in its important contribution to the overall record of our society. The problem we face in Canada is that without the intentional collection, systematic preservation, intellectual organization, and purposeful management of our research data, we will lose this part of our heritage.[28]

The challenge remains in Canada to form an institutional safety net for research data.

Endnotes

1. This paper draws upon research that was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) (Grant Nos. 421-2000-0011 and 421-2000-0017).

2. Shearer, K. *Preservation of Digital Resources: Selected and Annotated Reading List [with a] Primary Focus on Access Issues.* CARL/ABRC: December 2001. www.carl-abrc.ca/projects/preservation/reading_list.PDF

3. References in this paper to the National Library of Canada (NLC) and the National Archives of Canada (NAC) predate the amalgamation of these two institutions into Library and Archives Canada.

4. Green, A., J.A. Dionne, and M. Daniels. *Preserving the Whole: A Two-track Approach to Rescuing Social Science Data and Metadata.* Council on Library and Information Resources, 1999. www.clir.org/pubs/reports/reports.html

5. *Preliminary Draft Charter on the Preservation of the Digital Heritage* by the Director-General. www.unesco.org/general/eng/about/circulars/cl3643.pdf

6. The specific clause regarding research data is found in Article 10(c) of the *Revised Draft Charter on the Preservation of the Digital Heritage,* which was released in April 2003. portal.unesco.org/ci/ev.php?URL_ID=7869&URL_DO=DO_TOPIC&URL_SECTION=201&rel oad=1055384712

7. *Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program.* Washington, DC: Library of Congress, 2002. www.digitalpreservation.gov/ndiipp/repor/repor_plan.html

8. The National Data Archive Consultation was an initiative funded jointly by the NAC and the SSHRC. Beginning in the autumn of 2000, this working group investigated the need for national data archiving services in Canada as its initial task. In June 2001, a report was filed with the Council of the SSHRC and the National Archivist detailing the need for such services. Upon approval of this report, the working group undertook in the autumn of 2002 a second assignment in which it explored various models of providing data archiving services. The final report from this consultation was submitted to the Council of the SSHRC and National Archivist in June 2002. For further details see: www.sshrc.ca/web/whatsnew/initiatives/data_archives_e.asp

9. www.webopedia.com/TERM/d/data.html

10. Lyman, P., H.R. Varian, J. Dunn, A. Strygin, and K. Swearingen. *How Much Information?* University of California at Berkeley, CA: School of Information Management and Systems. www.sims.berkeley.edu/how-much-info/summary.html

11. *Ibid.*, p. 8.

12. For a description of the working group and its mandate see: Hackett, Y. "A National Data Archive for Canada?" *IASSIST Quarterly* 25, 3, pp. 13–16.

13. *National Data Archive Consultation: Phase One Needs Assessment Report.* Social Sciences and Humanities Research Council, June 2001. www.sshrc.ca/web/whatsnew/initiatives/da_phase1_e.pdf

14. Scientific data, which is a category of research data, should also be preserved to allow the replication of earlier findings from the data.

15. *Ibid.*, "2.0 Gaps in Existing Research Infrastructure," pp. 11–13.

16. For the United States see: Fienberg, S., M. Martin, and M. Straf (editors). *Sharing Research Data.* Washington, DC: National Academy Press, 1985. For the United Kingdom see: *Preserving and Sharing Statistical Material, Chapter 1.* University of Essex and the Royal Statistical Society: UK Data Archive, 2002. www.data-archive.ac.uk/home/PreservingSharing.pdf

17. The estimate in Canada is that the federal government spends more than $1 billion annually on the collection and management of data. See: *National Data Archive Consultation: Background Report.* www.sshrc.ca/web/whatsnew/initiatives/da_background_e.asp

18. Canada is one of the few G8 countries without national data archiving services. See: "2.2 Lost Opportunities" in *National Data Archive Consultation: Background Report.* www.sshrc.ca/web/whatsnew/initiatives/da_background_e.asp

19. *Data Policy and Barriers to Data Access in Canada: Issues in Global Change Research.* www.globalcentres.org/cgcp/english/html_documents/publications/data/toc.htm

20. Appendix I of the Phase I Report, pp. 21–35.

21. Humphrey, C. *A Survey of SSHRC-funded Researchers* (an internal report submitted to the NDAC in April 2001). The bulk of this report appears as Appendix I of the Phase I Report; missing is the appendix in the original report that contains the technical description of the sample.

22. This is not just a critical comment about researchers. There is an expectation for researchers to deposit data, but Canada does not have a recognized institution to receive data. One can easily understand why researchers would be confused.

23. Metropolis is a large-scale research program funded by the SSHRC and a consortium of federal departments and agencies. This is just one example of several other initiatives within the SSHRC through which research data are created outside SSHRC standard research grants.

24. Internal memorandum from John Saikkonen to Harold Naugler, Director MRA, entitled *Survey of SSHRC Grants, 7 November 1981. File 1600-1.*

25. The selection criterion used to identify funded research projects in this study was the presence of a budget item for computing resources. This criterion was particularly applicable during this study period because computing expenditures most likely covered the costs of mainframe computing cycles for the processing of data files or databases, although these costs could have also been to purchase a terminal to connect with a mainframe computer or to pay for manuscript processing. The nature of computing expenditures dramatically changed immediately following this period with the introduction of the IBM Personal Computer in 1981 and the subsequent explosion of network computing later in the decade.

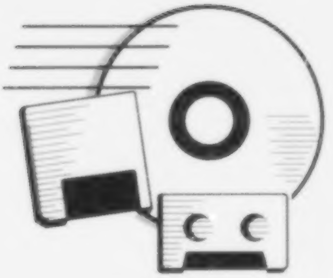
26. The following documents were consulted in determining the division between the social sciences and humanities:
Canada Council. *Annual Report 1977–1978.*
SSHRC. *Annual Report, 1978–1979.*
SSHRC. *Annual Report, 1979–1980.*
SSHRC. *Annual Report, 1980–1981.*

27. Included among the data catalogues that were consulted was CULDAT, the Canadian Union List of Data files developed by the MRA Division in the mid 1980s. This catalogue was the most comprehensive directory of Canadian research data that was available during the 1970s and early 1980s. The catalogue was not maintained following the 1986 reorganization in the NAC in which the MRA was disbanded. The three studies that were located with the ICPSR are the Canadian National Election Survey, 1979–1980, the British Columbia Election Study, 1979–1980, and the Quality of Canadian Life: Social Change in Canada, 1977.

28. Charles Humphrey in a letter written to John English on April 10, 1998 in a submission on behalf of the Canadian Association of Public Data Users to the Consultations on the Future Role of the NAC and the NLC.

Résumé

Les publications récentes portant sur la préservation des documents électroniques n'ont pratiquement pas tenu compte des « données de recherche » comme catégorie de produits numériques méritant qu'on s'y attarde. Cela s'explique notamment par la confusion générale qui règne sur la nature de ces données. La Consultation sur les archives nationales de données (CAND) au Canada fournit une définition des données de recherche qui distingue ces documents de tous les autres produits numériques. Cette communication étudie l'importance de cette définition et rend compte de quelques-unes des constatations de la CAND, qui fait le point sur le niveau de risque actuel que courent ces données. Une analyse des écarts a montré que ce type de données n'entre pas dans l'interprétation et l'exécution des mandats des institutions vouées au patrimoine canadien. En outre, il est prouvé que la perte de données peut devenir un problème grave lorsqu'une institution n'a pas le mandat précis de les préserver. Enfin, lorsqu'ils ne disposent pas d'une institution reconnue responsable de la préservation des données, les chercheurs ne savent comment ni où archiver leurs données de recherche.



ARCHIVAL SCIENCE AND ARCHIVAL ENGINEERING: BUILDING A NEW FUTURE FOR THE PAST

KENNETH THIBODEAU

Director, Electronic Records Archives Program
National Archives and Records Administration
College Park MD
USA
E-mail: ken.thibodeau@nara.gov

Abstract

Comparison of the knowledge and practice of archivists with those of engineers is particularly relevant at this time. There is an indisputable and very pressing need for archivists, and archival institutions, to engage in building systems and applications to enable successful management of electronic records from the moment of creation, and to ensure the possibility of preserving them for future generations. Archivists need to work with engineers, but how close can, or should, such collaboration become?

Introduction

At the National Archives and Records Administration (NARA) in the United States, we describe our efforts to respond to the challenges posed by electronic records as a project to build the archives of the future. This Electronic Records Archives (ERA) will not be a physical building, but a virtual repository in cyberspace. NARA's vision is that the ERA "will authentically preserve and provide access to any kind of electronic record, free from dependency on any specific hardware or software, enabling NARA to carry out its mission into the future."[1] ERA will be a set of capabilities that NARA organizations and staff, other government agencies, and the public can access from anywhere on the Internet to do their work, transact business, and find and retrieve records from NARA's holdings. It will be a system in terms of its coherence and comprehensiveness to support the management of records across their life cycle, but the location of the information technology infrastructure that enables this system will be, literally, immaterial to its users.

Through extensive market research, collaboration with computer scientists and engineers in research, and dialogue with large and small information technology companies, we have established that most of the basic technologies that are needed to build such a system are currently available as commercial products. However, even though the technologies exist, building a comprehensive system that is capable not only of preserving and providing sustained access to any type of electronic records, but also of supporting the process of managing records across their life cycles and interacting with innumerable other types of systems in federal agencies, researcher's schools, libraries, offices, and homes, and — not the least — in the future, is clearly beyond the state of the art. Numerous experts have characterized building the ERA as unprecedented, very complex, and extremely risky, but also very exciting because it promises to move the state of the art of information technology forward in a new direction — one that some analysts believe will be recognized as essential not only for cultural heritage institutions but also for conducting business in the digital arena. This undertaking is also very expensive. The President of the United States has requested US$36 million just to enable NARA to produce a viable architectural design for the system. For these reasons, this initiative is being very carefully watched, not only by NARA's management but also by the White House, the United States Congress, other government agencies, and the information technology industry.

One significant factor in the complexity of the system we want to build derives from NARA's decision, following sound principles of systems engineering, to create a system that will support NARA's entire end-to-end process of managing records across their life cycle. But given the fact that the basic process of managing records is independent of the types of records (and in fact requires an integrated approach that encompasses all of the records of any creator and also all of the records preserved in the archives), the scope of the system will have to extend to the life cycle management of all types of records. There is, nonetheless, one fundamental difference between the way the system will support the management of electronic records and how it will address other types of records. That is, the system will support the management of all records using data and information about those records, their creators, and the activities in which the records were created and used, but in addition the system will actually process electronic records. NARA has defined the processes needed to manage, preserve, and provide access to electronic records as an elaboration of the Open Archival Information System standard.

Building such a system is very challenging from a purely technological perspective, and the difficulty is greatly increased by the disparity between the way archivists think and work and the way that computer and communications systems are defined, designed, and developed. The problems are compounded further in the case of a national archives (or any archival institution that is responsible for preserving records created by a variety of persons or institutions) because the wide variety of sources of the records brings differences in both their technical and archival properties. The difficulties that NARA is facing in trying to build the archives of the future are not unique to NARA, but are common to many archives and other cultural institutions. Many of these problems are also faced by organizations that are not involved in cultural heritage, but whose business requires keeping information assets for long periods of time.

There is no doubt that we need to construct concepts, methods, and operational systems that can preserve and provide access to digital information while we are still at the dawn of the digital era, before too many cultural assets are lost and technology races utterly beyond our ability to catch up. Without concentrated efforts to build sustainable solutions for digital preservation, the present will have no future.

The need for collaboration between archivists and engineers

Let us consider what is involved in trying to build an infrastructure capable of preserving the records of our times, and of time to come. Given my organization and profession, I will concentrate on archives. The essential goal in preserving records is to enable reconstruction of past events, activities, decisions, and states of affairs from recorded information accumulated in the course of action. The role of archives and the archival profession is not to reconstruct the past — that is the proper work of historians, auditors, analysts, genealogists, and others who use records to extend their knowledge or to create new knowledge. The role of archivists and of archives is instead to deliver the building materials that others use to reconstruct the past. The systems that archivists must build for electronic records should not be mausoleums or reliquaries to house those records, but transportation systems to transport them, in authentic form, to the future.

In this role, we need to be aware of the burden of our own past, of the possibility that the theory, methodology, and practice that the archival profession has constructed in the past may hinder progress in developing or adopting optimal solutions for the digital era. That is not to say that we should jettison our professional heritage, but that we need to be judicious in carrying it forward; we must remain open to the possible need to deconstruct what we have already built to enable us to do our work in the future.

But archivists do not have the expertise required to develop technologies or to build systems that will satisfy archival requirements. However, this is an area where engineers may be able to help us.

In contrast to the natural sciences, history, or philosophy (where the goal is simply knowledge for its own sake), the ultimate goal for both engineers and archivists is the application of knowledge with practical, empirical results. Engineering obviously serves pragmatic purposes. Prominent engineer/historian Henry Petroski defines engineering as "the art of rearranging the forces and materials of nature for the good of society."[2] Likewise the archival profession applies its knowledge so that records are selected, protected, and provided to those who have an interest and right to them.

Although the archival profession, like engineering, works for the good of society, it parts company with engineers when it comes to "rearranging the forces and materials of nature." But that phrase is really too narrow with respect to both archives and engineering. With few exceptions, such as conservation laboratories and the occasional construction of archival buildings, the archival profession does not deal with the forces and materials of nature. And engineering is not always confined to dealing with the materials and forces of nature — it can be applied to other realms such as software engineering and financial engineering.

Another difficulty with "the art of rearranging ... forces and materials... for the good of society" is that "rearrange" has troublesome connotations for archivists. While some post-modernists might be comfortable describing archival practice as the art of rearranging or otherwise manipulating documentary materials for some supposed societal good, there is at least an apparent conflict with the principles of respecting the archival fonds and the original order of records. Even admitting the impact of social forces and subjective elements in the formation and management of archives, many archivists would see a basic conflict between their goal of transmitting to future generations the records of past activities and the common aim and outcome of engineering, which is the production of things that are new or different. Perhaps archivists and engineers can find common ground in one of the specialized areas of expertise that has been developed in systems engineering — the transition from "legacy systems" to new ones.

To ensure that the systems that are built satisfy records management and archival requirements, archivists will need to be able to communicate clearly with engineers and other technologists.

The development of the Electronic Records Archives

In its initiative to build the archives of the future, NARA is near to completing its articulation of the requirements the system must satisfy for NARA to do its business and accomplish its mission. Developing the requirements for such a complex system has been an extensive and intensive process.

NARA began by forming a series of Integrated Product Teams (IPTs), each of which brought together representatives of all the different stakeholders whom the system must serve. The ERA IPTs comprised specialists in all of the archival and records management functions assigned to the system, including electronic records specialists and also experts in textual, audio/visual, and cartographic records from the National Archives, Presidential Libraries, and Federal Records Centers around the country. Three IPTs progressively identified and refined the requirements, starting with a high-level concept of how the system will operate (described in narrative terms from the perspectives of the different types of users of the system) and ending with the articulation of more than 700 atomic statements of functional and system requirements.

Throughout this process, a fourth team (the Committee of Archival Requirements (CAR), that consisted of senior managers from NARA's line operations and was chaired by the Deputy Archivist of the United States) reviewed the products of the IPTs to resolve policy and other management issues that arose and validate that the final list of requirements will in fact express what NARA needs and wants to do to manage and preserve electronic records. For this purpose, the atomic statements of requirements — some 300 at the time — were translated into "use cases."[3] This was done by engineers and other technical staff, who ensured that the use cases developed reflected all identified functional requirements and highlighted situations where stepping through a process revealed gaps in the requirements. The ERA requirements were translated into 47 use cases that covered all the life cycle management processes within the scope of the system (including scheduling and appraisal, physical transfer of electronic records to NARA's physical custody, accessioning into legal custody, description, search, retrieval, and disposal) as well as generic system functions such as registering to use the system, monitoring system alerts, changing the user interface, and subscribing to notices on topics of interest. Processes were broken down into simple activities; for example, the process of developing disposition agreements was broken down into six separate cases such as creating a records schedule, submitting proposed records schedule to NARA, review within NARA, public comment, etc.

The CAR reviewed all of these use cases in its weekly meetings, often making significant revisions. The final versions of the use cases were then used to revise the existing ERA requirements, or to add new ones. The Requirements Document, revised to contain some 700 requirements, was released within NARA for comment as well as outside for public feedback in March 2003. All of the comments received — and there were several hundred specific comments — were used to further refine the Requirements Document. The revised version was released in August 2003, with a deadline of September 17 for comments on it.

To reach a successful conclusion, the final Requirements Document will need to articulate the requirements in a manner that adheres to both the criteria and constraints imposed by system engineering. This will require that they be simultaneously comprehensive and atomic, i.e. collectively they must describe everything that is needed but individually they must describe one and only one requirement.

To accomplish this, the requirements should initially be expressed in very broad terms so as to arrive at a limited set that all stakeholders can agree expresses everything that must be done. Subsequently, they must be refined to much greater precision in order to communicate to system designers clearly and unambiguously what is required. Once complete, the total set of requirements must not contain any redundancy which, especially in the high-level expression of requirements, can cause significant problems. For example, if redundant requirements get refined differently, it is possible to end up with inconsistencies or even incompatibilities at lower levels. Unfortunately, our experience indicates that getting archivists to express themselves in ways that satisfy the needs of systems engineering is rather difficult.

One of the difficulties is in expressing requirements in atomic, non-redundant terms. Archivists regularly deal with complex cases that have significant internal nuances, vary considerably from one case to the next, and have a lot of exceptions or special cases. As a result, the requirements for managing records are often expressed in terms that are broad, and even vague, in order to cover as much ground as possible; there is also a tendency to add a lot of qualifiers to ensure that exceptional cases and important subtleties are not excluded. This use of imprecise language can lead to problems, as indicated by the following example that surfaced only after several iterations of the Requirements Document. In the course of expressing requirements, archivists and records managers had described a variety of things that are done with records such as appraise them, describe them, preserve them, and so on. The IPTs went to considerable lengths to decompose these processes to yield clear descriptions of functional requirements. Technologists formulated the requirements statements from what the specialists in various areas of archives and records management told them, and the IPTs reviewed them. In the second version of the Requirements Document, however, it became evident that the technologists had not been sufficiently well informed to distinguish processes that apply to individual records from ones that operate on files or series. As a result, they interpreted the descriptions of life cycle processes as operating, in almost all cases, at the level of individual records. The records specialists clearly understood such differences, but they did not perceive the necessity of making things that are common knowledge within their discipline explicit when communicating to other disciplines. The IPT process makes eliminating such vagaries of language especially difficult, because it requires that the simple, clear statements that finally get recorded as requirements represent consensus of all the stakeholders represented on the team.

Another difficulty stems from the fact that, following sound principles of systems development as well as policy of the United States Government, requirements should be expressed in terms of what the institution and the users need rather than in terms of how the system will satisfy the requirements. That may sound simple, but in practice it is difficult. One of the obvious problems is prejudice, i.e. many people come to the table with preferences for specific solutions.

In fact, many people both inside and outside of NARA assumed that we intended to build an archival system for electronic records using technologies that had been demonstrated for us by the San Diego Supercomputer Center. But NARA's collaboration with San Diego never was, and never was intended to be, a system development or technology evaluation or selection. Instead, the partnership with the Supercomputer Center and other technology researchers was intended to expand the knowledge base that we could apply in building the archives of the future. Without an established solution that might be adequate for archival needs, we wanted to learn what might be possible given new technologies that were being driven as key elements of the next generation national information infrastructure. We needed to know what would remain impossible for the foreseeable future, and what might be feasible but would remain very difficult or risky. And we did learn valuable lessons. For example, both NARA and the computer scientists and engineers at San Diego learned that it is not possible to build an archival system that operates only at the levels of data management and information management. An archival system must also operate at the level of knowledge management. Archival science, archival concepts and principles, and even the lessons of archival experience have to be embedded in the system to address archival needs. But lessons of this sort do not prescribe particular solutions.

And it is important to exclude prejudice or preference for specific solutions from the articulation of requirements. For any complex system, it is impossible to know a priori if any particular solution — even if it is in fact the optimal way of satisfying some number of requirements — will be optimal given the total combination of requirements the system must satisfy.

One of the difficulties in keeping solutions out of requirements is that it can sometimes be difficult to distinguish between the two. For example, one of the electronic records experts working in one of the IPTs told the group that there is a requirement to produce tape maps.[4] Obviously that is a solution to the need to identify digital files, but it is not necessarily the only solution — it has just been in use for so long that it was considered necessary, i.e. a requirement.

Other cases are not so obvious. The requirements for ERA, for example, include the capability of materializing different file classification systems. This is needed because NARA must be able to accession electronic records maintained in Records Management Applications, which implement traditional filing systems. But is a file classification system a requirement for managing records?

This question is crucial to helping record creators develop applications that simultaneously enable them to do business and manage their records. Should we insist that information technology systems implement file classification systems? File classification systems are artifacts of the information technology of two-dimensional hard copy. When records are affixed to paper, the possibilities for organizing them are constrained: the most effective means of linking related records is by placing them in close proximity, and reinforcing this proximity by the physical boundaries of file folders and file drawers or shelves. The physical necessity is the root of filing systems, even though they are elaborated and communicated as conceptual arrangements. But are such filing systems really required to manage records? Just as digital technology is creating new types of records that could not exist without the technology, it opens up unlimited options for organizing records in support of the activities of the records creator. The ideal — and therefore the requirement — is that records be organized in a way that optimally supports the ongoing business of their creator, not that the order imposed on the records embody solutions that worked well on paper.

Consider the case of e-commerce or e-government, where an organization conducts business through a Web interface. The possibilities for arranging information on Web sites are myriad. Related documents do not need to be in proximity to one another — not physical proximity or even logical proximity. Not just documents, but specific elements of their content can be related to elements of other documents through hyperlinks. On the surface, then, the "arrangement" of records on a Web site can be very ad hoc. Below the surface, however, the structure may need to be regular and rigorous. To support the regular course of electronic transactions, the on-screen presentation of documents such as order forms, invoices, and status reports must be linked to underlying online transaction systems. This is the way information is organized to conduct business. To suggest that it should be arranged in a very different way, according to a traditional filing system, for purposes of record-keeping would introduce a layer of opacity between the user and the relationships that the records originally had to the processes in which they were created and used.

Another area where there is confusion between requirements and solutions is in discussions about techniques for digital preservation. Much of the recent discussion can be characterized as a debate on the relative merits of migration and emulation. Both of these methods are solutions. Trying to decide which is the superior method for preservation without specifying what requirements preservation must satisfy is like trying to determine what is the best medicine without taking into account the disease that is to be treated. Has the archival discipline clearly articulated the requirements for preserving electronic records in a manner that reflects a consensus of the profession — not to mention the stakeholders — and can unambiguously and comprehensively guide the design and development of systems that will satisfy those requirements?

This question leads us to shift perspective to the other communications problem described in recounting the development of requirements for the ERA system: the need for clarity, atomicity, and traceability in stating requirements. What should archivists communicate to systems developers concerning requirements that relate to the preservation of higher-level archival aggregates, such as series and archival fonds. There is a lack of unanimity and uniformity in the archival literature on these topics. The rich nuance in Peter Horsman's recent article on the "de-discovery" of the archival fonds, for example, reflects a state of affairs that is not readily amenable to interpretation in ways that can be implemented in systems.[5] One of the problems we will face in translating archival knowledge into systems specifications is that archival discourse has often been conducted at a high level of abstraction and generalization which is not readily traceable to clear, atomically expressed bases. This is an area where adoption of engineering methods for elaborating systems concepts could be of value to archivists. In order to appraise, describe, or communicate records, archivists need a deeper, more granular knowledge of how persons and organizations create and use electronic records as information assets in carrying out their activities.

Conclusion

In developing the archives of the future, we need to construct, or to reconstruct, a dynamic intellectual edifice. This new edifice will require pathways for systematically collecting and organizing empirical information about new types of electronic records and new ways that technology is applied in the conduct of business. It will need conceptual spaces suitable for analysing such information not only for its impact within the domain of the life cycle management of records, but also for what it reveals concerning changes in business processes and in the relationships between these processes and records. And finally it must have an infrastructure capable of integrating the results of such analysis into the solid foundations of archival science.

Endnotes

1. Era Vision Statement. 2002. www.archives.gov/electronic_records_archives/about_era.html#vision

2. Petroski, H. *Invention by Design: How Engineers get from Thought to Thing*. Cambridge, MA: Harvard University Press, 1996.

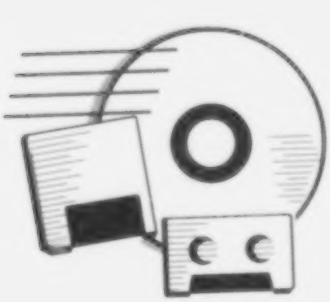
3. A use case is an object-oriented method for describing the processes executed by a system. Each use case describes the actors involved (either humans or other systems), the assumptions and preconditions of the case, the steps taken in carrying out a process (including alternate scenarios as appropriate), the results of the activity, post-conditions, and any issues identified in analysing the case.

4. A tape map is a tool that archivists have used for decades to determine exactly what is on media when transfers of electronic records are received.

5. Horsman, P. "The Last Dance of the Phoenix, or the De-discovery of the Archival Fonds." *Archivaria* 54 (2002), pp. 1–23.

Résumé

La comparaison des connaissances et des pratiques des archivistes et des ingénieurs est particulièrement utile à l'heure actuelle. Les archivistes et les services d'archivage ont absolument et instamment besoin de se mettre à élaborer des systèmes et des applications qui leur permettront de gérer leurs documents électroniques dès l'étape de la création et de veiller à ce qu'ils puissent être préservés pour les générations à venir. Les archivistes doivent travailler avec des ingénieurs, mais cette collaboration doit-elle être étroite et dans quelle mesure doit-elle l'être?



DIGITAL PRESERVATION STRATEGIES: THE INITIAL OUTCOMES OF THE ERPANET CASE STUDIES[1]

SEAMUS ROSS

Director, Electronic Resource Preservation and Access Network (ERPANET)
Humanities Advanced Technology and Information Institute (HATII)
University of Glasgow
Glasgow, Scotland
United Kingdom
E-mail: s.ross:@hatii.arts.gla.ac.uk

MONICA GREENAN

British Content Editor ERPANET
(May 2002–April 2003)
HATII
University of Glasgow
Glasgow, Scotland
United Kingdom
E-mail: british.editor@erpanet.org

PETER MCKINNEY

Coordinator ERPANET
HATII
University of Glasgow
Glasgow, Scotland
United Kingdom
E-mail: p.mckinney@hatii.arts.gla.ac.uk

Abstract

The Electronic Resource Preservation and Access Network (ERPANET) is a key European Commission Fifth Framework digital preservation activity. Among its other activities, it has recently conducted a series of case studies to improve its knowledge of digital preservation practices within Europe's public institutions and private-sector companies. Information for the case studies was collected between August and November 2002 by interviewing staff from 20 companies spread across the broadcasting, pharmaceutical, publishing, and telecommunications sectors. This paper describes the methods used to conduct this research and presents the results; it also defines areas where further work (e.g. research, guidelines, improved practices) is urgently needed.

Introduction

Records managers, archivists, and librarians all recognize the need for urgent action to ensure the survival of digital information. In order to understand how European organizations are approaching digital preservation, ERPANET[2] has been conducting case studies to investigate awareness, strategies, and technologies. More than 500 organizations, institutions, and public bodies will eventually contribute to this research — even if for some it may only be by taking the decision not to take part.[3] The results should make a substantial contribution to current knowledge of practice in digital preservation, and form the foundation for theory-building and the development of methodological tools. The value of these case studies will come not only from the breadth of sectors included, but also through the depth at which the issues will be explored.

ERPANET is deliberately and systematically approaching disparate organizations from industry and business to facilitate discussion in areas that have traditionally been unconnected. With these case studies, ERPANET will be able to broaden the scope and understanding of digital preservation through research and discussion. The case studies will be published online on the ERPANET Web site[4] to help improve the approaches and solutions being developed and to reduce the redundancy of effort. The interviews are identifying current practice not only in-depth within specific sectors, but also cross-sectorally, for example: what can the publishing sector learn from the aeronautical sector? Eventually we aim to use this comparative data to produce cross-sectoral overviews. This cross-sectoral fertilization is a main focus of ERPANET as laid out in its Digital Preservation Charter.[5]

Aims

The principal aims of the case studies are to:

- build a picture of methods and match them against context to produce best practices

- accumulate and make accessible information about practices
- identify issues for further research
- enable cross-sectoral practice comparisons
- enable the development of assessment tools, through the identification of tools that are needed and the accumulation of information about good practices
- create material for training seminars and workshops
- develop contacts

Methodology

Potential sectors on which to base the case studies are selected to represent a wide scope of information production and digital preservation activity. Each sector presents a unique perspective on digital preservation, as organizational and sectoral requirements, awareness of digital preservation, resources available, and the nature of the digital object created place unique and specific demands on organizations. Initial desk-based sectoral analysis provides ERPANET researchers with essential background knowledge. The primary research is then conducted by interview.

In developing the interview instrument, the project directors and editors reviewed other projects that had used interviews to accumulate evidence on issues related to digital preservation. Among these, the methodologies of the Pittsburgh Project[6] and InterPARES I[7] for target selection and data collection were given special attention. But after careful consideration, the Pittsburgh approach was considered too narrowly focused and provided insufficient breadth to enable sectoral comparisons. And the InterPARES I data collection methodology proved too detailed and lengthy (which could become an obstacle both at the point of data collection and during analysis) and focused too closely on record-keeping systems (we wished to take a wider view).

The ERPANET interview instrument that was eventually developed (see Appendix[8]) takes into account both the strengths and weaknesses of earlier projects, and targets a range of strategic points in the organizations under examination. It explores three main areas of enquiry within an organization: awareness of digital preservation and the issues surrounding it; digital preservation strategies (both in planning and in practice); and future preservation requirements. Within these three themes, distinct layers of questions uncover the state of the entire digital preservation process within the institution.

Drawing on the experience that the partners of ERPANET have in this method of research, another feature is also included, i.e. wherever possible three categories of employee within each organization are interviewed: an Information Systems or Technology Manager, a Business Manager, and an Archivist/Records Manager.[9] This provides two members of staff with knowledge of the organization's digital preservation activities, and a manager to offer an overview of business and organizational issues. When completed, this interview instrument provides information on the extent of knowledge and practice in organizations, who has responsibility and problem ownership, and where the drive towards digital preservation is initiated within organizations.

The main areas of investigation include:

- perception and awareness of risk associated with information loss
- level of understanding of how digital preservation affects the organization
- actions that have been taken to prevent data loss
- processes for monitoring preservation activity
- mechanisms for determining future requirements

Within each section, the questions are designed to bring organizational perceptions and practices into focus. Questions are aimed at understanding impressions held on digital preservation and the impact that it has on the respective organization, exploring the awareness in the sector of the issues and the importance that digital preservation is accorded, and determining how digital preservation affects organizational thinking. For instance, participants are asked to describe (in their views) what the main problems associated with digital preservation are and what long-term value information actually has in their sector. Through this the reasons for preserving information as well as the risks associated with not preserving it become clear.

The core of the questionnaire focuses on the actions taken at corporate and sectoral levels in order to uncover policies, strategies, and standards currently employed to tackle digital preservation concerns, including selection, preservation techniques, storage, access, and costs. Questions allow participants to explore the future commitment from their organization and sector to digital preservation activities and, where possible, to relate their existing or planned activities to those being conducted in other organizations with which they might be familiar.

The first round of the case studies focused on companies drawn from the pharmaceutical, broadcasting, publishing, and telecommunications sectors. These sectors were chosen to guarantee a diversity of organizational type, activity, regulatory framework, and culture. The aim was also to include a range of institutional types, sizes, and locations in the case studies, although this was not always possible (e.g. in the telecommunications sector).

This survey process was not without its problems. Even when organizations were identified and interviews time-tabled, targets often withdrew just before we began the interview process. Some withdrew after seeing the data collection instrument and realizing the time/effort involved, and others (we suspect) because they realized that the expertise was not available within their organization to answer the questions. In the latter instance, withdrawal might indicate that the organization was not maintaining their digital records effectively. The perception of risks that might arise through contributing to the study worried some organizations, particularly those from sectors where competitive advantage is imperative, or liability and litigation issues especially worrying.[10] Telecommunications proved to be one of the most difficult. Non-disclosure agreements that stipulated that we would neither name an organization nor disclose any information that would enable readers to identify them without their explicit approval were used to reduce risks associated with contributing to this study. In some cases the risk was still deemed too great and organizations withdrew. Several organizations withdrew on seeing the final report because its conclusions were unsettling (one particularly worrying instance from a later set of case studies was the withdrawal of a major national nuclear energy producer on sight of their report).

Overview of findings from the first round of studies

This first round of case studies was undertaken between August and November 2002. Nearly 50 staff[11] from 20 organizations in the pharmaceutical,[12] broadcasting,[13] publishing,[14] and telecommunications[15] sectors were interviewed. Findings from the case studies showed wide variation in current strategies and practices. The organizations surveyed had a huge amount of digital information to handle, and the types of digital objects themselves varied across the sectors. A wide variety of reasons and justifications for maintaining this information existed. Diversity in preservation activity reflected the different levels of understanding of the issues and a lack of consensus on the best methods to meet preservation challenges.

Perception and awareness

An assessment of the levels of awareness of both digital preservation issues and the value of information was included to gauge the depth of understanding and the relevance of long-term access to digital materials to the individual organizations. Uncovering what preservation issues were most prominent for both the organizations and the sectors provided a mechanism to assess the relevance of digital preservation for them.

The two sectors with the deepest awareness were the pharmaceutical and broadcasting sectors, with the broadcasting sector in particular displaying an acute understanding of the challenges of digital preservation. This is a sector that actively pursues solutions collaboratively. This awareness reflects their widespread appreciation of the re-use value of broadcasting materials. The pharmaceutical sector has had digital preservation awareness imposed by external regulators, who have created a complex and challenging structure of compliance and an economic advantage for the filing of drug approval requests in digital form. The driving force behind the development of solutions by the pharmaceutical sector is the legal requirement to preserve information relating to many of its products, although the fiercely competitive nature of the individual organizations within the sector has resulted in these solutions being developed in isolation. Similarly, the preservation activities of the telecommunications sector are driven by legal requirement rather than the desire to generate historical records. Thus data on call records, contracts, and interconnect data are kept for only 10 years. The publishing sector has become increasingly aware of the problem in the last few years, especially with the growth in electronic journal publication. Publishers are still attempting to establish who is responsible for the preservation — the publishers or the libraries (which have, in the print environment, traditionally played a central role in this activity).

It is not really surprising that the values attributed to information by organizations and sectors varied. Where most of the organizations fell down, however, was in their failure to transform their digital material into assets.[16] Respondents typically classified information as either having a re-use value or a cultural or historical value. However, only the most established publishing organizations and some of the broadcasting organizations were concerned about preserving information for its cultural or historical value; a substantial amount of what is produced in these sectors is considered a cultural asset. In the pharmaceutical sector, there were two clear opinions on the re-use value of information: those who believed that scientific research information can be re-used (the expense of creating research data and the vast amounts created were the main reasons for this attitude); and those who felt that if re-investigation had to be carried out in the future, technological advances would make the existing data obsolete. Within the publishing sector, some of the older publishers were concerned about the historical value of information, but for most the re-use value was the most prominent issue: access to publications was essential to enable their republication and to create multiple ways to exploit the products. However, within the publishing sector, individuals often attributed value to digital materials in the absence of a clear organizational position.

Cooperation

By including a focus on knowledge sharing, the case studies aimed to discover current and future levels of cooperation.

Results showed that in most cases collaborative effort to tackle the problems of digital preservation was rare. However, within the broadcasting sector, a lot of work has been undertaken to develop standards and best practices for the management and preservation of broadcasting material. International organizations such as the European Broadcasting Union (EBU)[17] and the European Commission have funded research projects such as the Preservation Technology for European Broadcast Archives (PRESTO)[18] which have fostered the development of strategies for the preservation of film, video, and audio material. Efforts have also been channeled into the exchange of information, cooperation on shared projects, and professional training. There is some discussion among the pharmaceutical sector but a specific focus on digital preservation is rare — and when discussions do take place there is little in the way of detail due to the intense competition inherent in this sector. Pharmaceuticals expend large amounts of resources on tackling preservation problems by themselves without much concern for the techniques and strategies developed by competitors or companies from other sectors. Common initiatives are also unusual in the publishing sector, but publishers are at least aware of what is going on in other organizations. Many report that they anticipate collaborative work with libraries as the next likely development; the collaboration between the National Library of the Netherlands (Koninklijke Bibliotheek) and Reed Elsevier is an example of the potential here.[19]

Policy, strategy, and standards development

In identifying current practice within organizations, the case study attempted to uncover the variety of policies, strategies, and standards that have been developed, adopted, and implemented.

All of those interviewed were aware of the need to identify and implement solutions, and develop them.[20] For those that had developed specific strategies and policies, these were often not implemented across all departments and satellites.[21] Broadcasters indicated that they exchange information on costing policies, strategies to develop technical solutions, standards, and implementation guidelines. This effort is reflected internally with directives, standard procedures, and preservation programs dealing with preservation requirements, recovery, standard format, and metadata systems. For other sectors, policies are developed internally with additional support from external service providers. Most pharmaceutical organizations reported that they plan to develop preservation policies as part of their projects articulating system requirements and specifications to software suppliers and system developers. Among our cohort were some who reported that their organizations do not plan to adhere to standards and guidelines established by other communities, although they do intend to review other policies and standards and build on them. Publishers have few strategies in place, and this may reflect a lack of systematic approaches to preservation. Most indicated that a good strategy was to keep everything; this, in their view, at least ensured that the material would be there in the future. But the approach begs questions about documentation, formats, and curation. There is a need, therefore, for a general raising of awareness of appraisal strategies.

Selection of digital information for preservation

The results indicate that most staff members have a solid theoretical awareness of the issues and the need to determine retention. This does not, however, mean that this knowledge is being put into practice. Within the publishing sector, if there is a need to keep information, this is done within an organization on an individual or departmental basis. Often it is those with a stake in the information and its re-use value who act to secure it rather than the organization itself. This ad hoc practice has operated successfully in the past; hence there is little impetus for the adoption of more formal controls. Pharmaceutical organizations preserve all classes of information that regulating authorities stipulate must be retained. In this case, selection criteria from the paper world have been transferred to the digital. As a result of falling storage costs, many organizations do not consider it worth the effort to filter digital information but instead retain all data. Selection criteria in the broadcasting sector take into account archival, technological, and practical considerations, with particular emphasis on re-use value. Internal directives that also set classification and management procedures of retention schedules regulate selection criteria. Only a few organizations indicated that any sort of selection policy had been developed in consultation with internal departments and organizational units with a stake in the preservation of digital objects.

Preservation and storage

Within the pharmaceutical environment, the size and proliferation of formats are the main obstacles to the preservation of objects. Overall the interviewees indicated that their organizations show little impetus to develop solutions to ensure digital longevity: the majority believe that solutions should be developed by external, specialized suppliers in order to allow organizations to concentrate on their actual field of operations and area of expertise. A pragmatic approach is the one most generally taken, with the migration to new data formats undertaken when the need arises, although few organizations seemed to be aware of the complexities associated with migration. To be safe, some organizations retain the digital object in its original format alongside the migrated version. In regulated sectors, such as pharmaceuticals, authenticity, integrity, and confidentiality of the records must be guaranteed. Validating these elements at the level of the individual digital object would be prohibitively expensive. The optimum point of validation is system level, with support of transaction audit trails. For the digital objects themselves, pdf (portable data format) has emerged as the prevalent preservation format, with several organizations restricting the use of its special features to enhance its effectiveness. A technique for the long-term preservation of more complicated or specialized data formats remains to be adequately addressed. The effort involved in convincing suppliers of scientific equipment to use standard formats was seen as too great due the time that would need to be spent on developing such agreements. Of the four sectors, it is perhaps the pharmaceutical sector that has sufficient power to promote a change in attitude on the part of suppliers. Although publishing and broadcasting together exercise greater financial muscle (as measured by annual turnover) than the pharmaceutical sector, issues of liability and risk do not play as significant a role for them.

Responsibility for preservation activity in the publishing organizations falls on the information technology departments. Results indicated that pdf is a very popular format for distribution and preservation; other formats include tiff (Tagged Image File Format), xml (Extensible Markup Language), and sgml (Standardized General Markup Language). Most organizations noted that they are waiting for an industry format to emerge. Preservation planning is made either as an ad hoc decision or because of a vendor relationship. Some organizations still consider retention of a print or microform copy as the most reliable preservation option currently (2002) available.

The participating telecommunications organizations reported that their sector integrates preservation activities into their business software, and migrations are conducted with every update to newer versions of the software. As 10 years is the maximum retention period, this approach is considered sufficient. But no standard metadata sets were applied, and there was little appreciation of their value. One organization interviewed did transfer expired files onto CD-R for preservation, but only as a matter of historical interest — and no resources were expended on monitoring the material or maintaining its accessibility after the initial transfer.

For broadcasters, results indicated that preservation activities are either carried out entirely in-house or partially outsourced as a way of reducing costs, and training of personnel is considered fundamental to the process. Most of the material produced for programming is converted to digital format, if not already in that state. Metadata production and collection is a central part of the access and preservation system, and a substantial investment of human and financial resources is deployed in this area. The formats currently used by this sector are considered reliable and migration is not seen as an immediate issue. Out of all the sectors explored, broadcasting sees the greatest need for the long-term preservation of its materials; one participant suggested that an international repository that would take responsibility for digital materials should be set up.

Digital preservation costs

The cost of digital preservation concerned all organizations interviewed. Some see the financing of digital preservation activities as corporate investment, whereas others pass responsibility to research and development divisions. This dichotomy exists because of the division of responsibilities. In a few organizations the responsibility for preservation is shared by all parts of the organization, and hence the costs are spread across the stakeholders. In these organizations, viable and manageable cost mechanisms usually exist.

The broadcasting sector expends substantial resources on digital preservation, and is now trying to streamline activities in order to rationalize its spending. They will concentrate on continuing digitization efforts, providing more access (e-delivery) and developing mass storage solutions. Where there is a specific budget for digital preservation, it commonly amounts to 1% of the organization's total budget. Allocations are identified through an accurate cost-and-benefit analysis based on previous experience and the results of various initiatives. Occasionally broadcasters are able to secure additional external funding because their digital objects are considered to be cultural assets. The availability of European Commission funding has been a major factor in fostering cooperation among broadcasters.[22] Interviewees from the publishing sector reported that their companies would like to spend more on digital preservation, but recognize that a better understanding of the costs involved has to be established before any investment will be possible. In order to increase spending, the budget holders must be convinced by a solid business case. Pharmaceutical organizations were concerned by the risks involved in spending on digital preservation, but felt that the current level of expenditure was broadly right. A few wished to spend more on preserving the vast quantities of information resulting from research and development activity for its research potential.

Monitoring of actions

With strategies and techniques in place, auditing procedures offer crucial mechanisms to measure compliance over time. Compliance with legal requirements and, where in existence, organizational directives, were common ways of monitoring activity. In more than 50% of the target organizations there was little or no monitoring. This reflected the lack of formalized preservation processes in the first place. Aside from internal monitoring, pharmaceutical organizations are subject to external inspections by regulatory agencies to audit their compliance. It is compliance and risk that dictate the processes that the pharmaceutical sector has undertaken, and is planning to undertake. Less regulated sectors have not been as quick to address the preservation problems.

Future requirements

Most participating institutions indicated that they believe their current state of activity is sufficient in the short term, and they are confident that their organization will be able to deal with any problems that arise in a manner such that information will not be lost and the organization will not suffer any detrimental effects, including financial loss or increased risk. To ensure that this will indeed be true, the need to preserve an increasingly large quantity of records and information should be linked to a business case to improve and expand access to it. Some sectors reported that solutions in such areas as mass storage and migration would be best procured from external commercial suppliers. Both the publishing and the telecommunications sectors were keen to see increased development of clear and implementable organizational strategy and policy, with commitment and support from the senior levels of management. The publishing sector in particular indicated their desire to increase the number and competencies of personnel dedicated to digital preservation within organizations. Pharmaceutical organizations would like to see the disparate preservation activities and storage brought together in an organization-wide solution.

Conclusions

This paper has drawn attention to the diverse nature of digital preservation strategies and practices. It has also examined the differences in requirements between sectors, but despite these differences all institutions would still benefit from the intra- and cross-sectoral sharing of knowledge. Most of the organizations, with the exception of the broadcasting sector, are hesitant in their digital preservation activities. They are waiting for external developments that they can adopt, or off-the-shelf solutions that they can implement. Common responses indicate that, due to the fragmentary nature of most of these organizations, little is undertaken on an organizational basis; instead, responsibility for preservation is shifted down either to department level or, in many cases, the individual.

Hesitancy arises from a lack of awareness in some sectors. Not only in terms of the enormity of the problem, but also in terms of solutions that have been developed in other areas. Indeed, one of the respondents indicated that when first approached by ERPANET about participating, the initial contact e-mail passed through the hands of ten people before someone was found in the organization who understood the problem well enough to get involved.

The broadcasting example indicates that strong inter-institutional cooperation leads to increased confidence and, in turn, increased activity and development of common knowledge and understanding. Furthermore, it is widely regarded that cooperation will lead to the development of more robust and financially viable tools and solutions. The pharmaceutical sector might learn from the success of the broadcasting sector, but whether their inherent risks in sharing information would outweigh any benefit would need to be assessed. The successes in the telecommunications sector, where actions and responsibility are spread around the organization, offer a streamlined model that others would benefit from implementing.

The case studies to date have outlined the knowledge, requirements, skills, and practices in digital preservation. However, communication and a common platform for information sharing are also key to the success of individual organizations' strategies. Recommendations from this round of studies are that intra-sectoral collaboration take place and that cross-sectoral discussion be encouraged. There are definitely areas where sharing can occur — strategies, costing models, and general awareness. From these case studies ERPANET has identified five areas where it will promote further action:

- creation of off-the-shelf policy statements
- development of business cases and strategies
- drafting of clear guidance on technologies and their preservation implications
- constructions of improved models (reference, costs, standards, functional requirements)
- production of guidance on creating, managing, and auditing digital repositories

It is with this in mind that ERPANET began in mid 2003 to develop briefing papers laying out guidelines that will enable organizations to measure their "preservation effectiveness" and improve their digital preservation practices. We are also beginning to engage in discussion with software vendors aimed at convincing them of the importance of incorporating preservation functionality into their products. Through these activities we hope to promote the development of cross-sectoral strategies for digital preservation.

Endnotes

1. The authors would like to thank their ERPANET colleagues Andreas Aschenbrenner (Nationaal Archief Nederland, Den Haag), Georg Büchler (Schweizerisches Bundesarchiv), and Prisca Giordani and Vincenzo Di Meo (ISTBAL, Università di Urbino) for helping to conduct the case studies.

2. ERPANET is a European Commission funded activity (IST-2001-32706).The Directors of ERPANET are Niklaus Bütikofer (Schweizerisches Bundesarchiv), Maria Guercio (ISTBAL, Università di Urbino), Hans Hofman (Nationaal Archief Nederland, Den Haag), and Seamus Ross (HATII, University of Glasgow). See the ERPANET Web site (www.erpanet.org) for more details and available products.

3. Since the original paper was submitted, ERPANET has decided to conduct a broader survey of leading companies in the 15 EU Member States and the 10 countries that are about to join the EU. See the ERPANET Web site (www.erpanet.org) for more information.

4. Case studies are made available on the ERPANET Web site (www.erpanet.org) as soon as they are completed.

5. The Charter is ERPANET's statement on the principles of digital preservation. It has been drafted to achieve a concerted and coordinated effort in the area of digital preservation by all organizations and individuals that have an interest in and share these concerns. It can be viewed on the ERPANET Web site (www.erpanet.org/www/content/documents/Digitalpreservationcharterv4_1.pdf).

6. The Web site of the Pittsburgh Project was lost due to a "technical glitch." However, interested individuals are encouraged to go to the Internet Archive site (www.archive.org/) and use the "Wayback Machine" to find the Pittsburgh Project (www.sis.pitt.edu/~nhprc). Our evaluation was carried out using a hard copy of the materials produced by the Pittsburgh Project which was provided to Seamus Ross in 1995 for his review *Commentary on the Pittsburgh University Recordkeeping Requirements Project: A Progress Report (Delivery Draft)*, Society of American Archivists, 59th Annual Meeting, Washington, DC (Thursday 31 August 1995, Session 20); this review is available at ERPAePRINTS (eprints.erpanet.org/).

7. www.interpares.org

8. We include the interview instrument to encourage comment and in the hope that others conducting similar research might adopt it (or a variant of it) as a way of increasing the comparability between different studies.

9. See the individual case study reports (www.erpanet.org) as this was not always possible.

10. It has been suggested that ERPANET should present the results of the studies as anonymous reports (e.g. An International French Bank, a European Nuclear Generator), because the participation rate might increase if companies knew they would not be specifically named. The Directors considered this approach, but concluded that case studies with identifiable targets have authenticity and "realism" that anonymous reports lack.

11. The original target of three staff per organization was not always achieved; see the individual case study reports (www.erpanet.org) for an explanation.

12. Astra Zeneca PLC (www.astrazeneca.com/); Aventis Pharma Germany GmbH (www.pharma.aventis.de); Bayer AG (www.bayer-pharmaceuticals.com); Böhringer Ingelheim (www.boehringer-ingelheim.com); GlaxoSmithKline plc (GSK) (www.gsk.com); Organon NV (www.organon.nl/); Pfizer (www.pfizer.com). Fourteen other pharmaceuticals were contacted and, although a number expressed initial interest in taking part, they all eventually decided not to contribute to the study.

13. RAI (Radiotelevisione Italiana) (www.rai.it); BBC (British Broadcasting Corporation) (www.bbc.co.uk); ERT (Elliniki Radiophonia Tileorassi SA) (www.ert.gr); SF DRS (Schweizer Fernsehen der Deutschen und Rätoromanischen Schweiz) (www.sfdrs.ch).

14. Gruppo Mondadori (www.monadori.com); Hachette Filipacchi Médias (www.hachette-filipacchi.com); HMSO (Her Majesty's Stationary Office) (www.hmso.gov.uk); Kluwer Academic Publishing (www.wkap.nl/); Office for Official Publications of the European Communities (Publications Office) (eur-op.eu.int/general/en/ index_en.htm); Oxford University Press (www.oup.co.uk/); Random House (www.randomhouse.com/home.pperl).

15. Swisscom (www.swisscom.com); Orange (www.orange.com). Numerous other telecommunications companies across Europe were approached, but most refused to take part or dropped out. When preparing this paper we considered dropping the telecommunications sector because it included too few companies, had a poor geographic spread, and did not include enough variation in organizational size. However, because the results from the telecommunications sector were consistent with the results from other sectors, we decided that it would be beneficial to include them.

16. For more on this see: Moon, M. "How do Cultural Artefacts become Digital Assets?" *Digital Asset Management Systems for the Cultural and Scientific Heritage Sector. DigiCULT Thematic Issue 2* (2002).

17. www.ebu.ch/

18. presto.joanneum.ac.at/index.asp

19. www.kb.nl/kb/resources/frameset_kb.html?/kb/pr/pers/pers2002/elsevier-en.html

20. The ERPANET Seminar on Policies and Procedures held in Fontainebleau in January 2003 added further weight to the findings of these case studies. There was a demand for guidance in the development of policies, and even those organizations that had them in place acknowledged that they were not always implemented. For more information see the ERPANET Web site (www.erpanet.org/www/products/Paris/paris.htm).

21. Even where policies are in place they are often unrealistic and, as a result, unimplementable.

22. Projects such as COLLATE (Collaboratory for Annotation, Indexing, and Retrieval of Digitized Historical Archive Material) (www.collate.de), AMICITIA (Asset Management Integration of Cultural heritage In The Interexchange between Archives) (www.amicitia-project.de/index.shtml), PRESTO (Preservation Technology) (presto.joanneum.ac.at/index.asp), and FIRST (Film Restoration and Conservation Strategies) (www.film-first.org/emf/first.asp?cod_seccao=208) were made possible under the IST (Information Society Technology) Fifth Framework Programme.

Appendix: ERPANET case study data collection instrument

Administrative Section
Interview Details
Organisation Details
Disclosure/Privacy Information
Tracking of Activities

Perception and Awareness of Digital Preservation

We would like to begin by asking you a few questions about your general impressions of digital preservation, and the impact that it has on the __________ sector. We will use the term 'digital information' throughout to refer to all forms of digital data, records and information.

1. Is there a general awareness in the __________ sector that the long-term preservation (more than five years) of digital information is an important issue?
2. To what extent does the sector recognise the importance of preserving digital information in the long term?
3. What are the main problems associated with digital preservation in the __________ sector?
4. From what sources have you heard about the issues surrounding digital preservation?
5. What values does digital information have in the ______________sector beyond the original purposes for which it was created?

Understanding How Digital Preservation Affects Your Organisation

We would like to focus on how some of these digital preservation issues affect your own organisation.

6. What type of information is digitally preserved in the short and the long term in your organisation?
7. What are the reasons that digital information is preserved in your organisation?
   ☐ Legal requirements
   ☐ Financial requirements
   ☐ Business requirements
   (e.g. document important decisions and activities)
   ☐ Historical value
   ☐ Other (please specify)
8. What risks is your organisation under if digital information is not preserved in the long term?
   ☐ Legal risks
   ☐ Financial risks
   ☐ Business risks
   ☐ Historical value
   ☐ Other (please specify)
9. Has the organisation conducted a risk analysis and/or business needs analysis with regard to the preservation of information? If yes, can you indicate the main results?

Actions Taken: Policies, Strategies, Standards and Practices Developed

The questions in this section aim to explore some of the actions that the organisation has undertaken to deal with the preservation of electronic records. It will examine the above as well as selection, preservation, storage, and access activities.

Policies, Strategies, and Standards

10. Is there any collaborative effort across the __________ sector to tackle common digital preservation issues?
    ☐ Conferences
    ☐ Newsletters
    ☐ Journals
    ☐ Common institutions
    ☐ Collaborative projects
    ☐ Other (please specify)
11. Has your organisation attempted to find information external to the sector regarding preservation?
    If yes, please indicate the sources.
    ☐ Government agencies
    ☐ Higher education institutions
    ☐ Archives
    ☐ Libraries
    ☐ Museums
    ☐ IT specialists
    ☐ Other (please specify)

    Please specify the kind of information provided and how useful it proved to be.

    ____
    ____
    ____
    ____
    ____

12. Do you cooperate with other institutions in the research and development of policies, strategies, and standards? In what way?

13. How useful is this common effort in applying it to your organisation's own needs?

14. Do you have any specific organisational policies that relate to the preservation of information?

15. Who (and what) was/is involved in the creation of these policies?
☐ Management
☐ Employees
☐ Special task force in the organisation
☐ Results of internal analyses (e.g. risk analysis)
☐ External sources, models, advice
☐ Other (please specify)

16. Do these policies apply across the entire organisation?

17. How are these policies implemented?

18. Has your organisation developed preservation strategies, standards, and practices and implemented them?
☐ Yes
☐ No
If yes, please specify.

19. How were they introduced and implemented (e.g. by department, with training)?

20. How, and under whose responsibility have these been established?
☐ External advice/sources/models
☐ Survey of information resources
☐ In-house solutions developed
☐ Other (please specify)

21. How often are your preservation policies and strategies updated and renewed?

• Selection of Digital Information for Preservation

22. Do you have a selection policy, or classification and retention policy that determines what information in your organisation is to be preserved?
☐ Yes
☐ No
If yes, please specify.

23. Is your classification and retention schedule linked and implemented across the organisation?

24. Who is responsible for the maintenance and implementation of these schedules?

25. How do you ensure that selected information is complete, accurate and identifiable?

Preservation of Digital Information

26. Does your organisation take care of its preservation activities itself, or are these outsourced?
☐ Outsourced
☐ In-house
If outsourced, what reasons were behind this decision, and who carries out the preservation activities?

27. Are there specific individuals in your organisation responsible for the preservation of digital information?

28. What positions do these people hold in the organisation, and what are their responsibilities and competencies?

29. What type of training or advice is available for them?

30. Is your organisation aware of any external standards, best practices, and guidelines available on preservation?
☐ Yes
☐ No
If yes, please specify.

31. Are these specific to your sector?

32. Where did you learn about them? Please specify your sources.

33. Which of these standards, practices and guidelines do you use?

34. What technologies do you use for preservation? For the following list of current techniques, please specify which ones you use and for what kind of information.

Technique	Specify type/ technology used	Information preserved
Print to paper		
Scanning		
Save on disk		
Save on other media		
Emulation		
Migration		
Microfilm / microfiche		
Other		

35. On what grounds were these techniques chosen? Please specify your answers.

☐ External advice
☐ Trials and evaluations
☐ Recommendations
☐ Intra-sectoral standards available
☐ Other

Please provide as much information as possible about why these decisions were taken.

36. What data formats do you use for preservation?

☐ Standard data formats
☐ Other

Please specify for both answers.

37. Do you convert the information to be preserved into other data formats for technical (or other) reasons?

38. What metadata do you use to describe both your digital information and the processes of storage and preservation? Does it follow any standards available (Dublin Core or other)? Can you provide a copy of the metadata set?

39. Is the collection and production of metadata automated?

40. Who is responsible for the transfer of information into long-term storage?

41. How often (if undertaken) is digital information migrated or refreshed?

Storage of Digital Information

42. Do you have a particular storage area for digital information to be preserved?

☐ Yes
☐ No

If yes, how is this organised and equipped?

43. Do you keep redundant copies of the digital information to be preserved for safety (or other) reasons?

Access to Digital Information

44. How is information protected from inadvertent or unauthorised access and manipulation?

45. Does your preservation solution allow direct access to the digital information stored (i.e. are they stored in an executable format)? If no, how is the access provided?

46. What access issues does your organisation face?

a. Copyright
b. Privacy issues
c. Access security and privileges
d. Other (please specify)

47. How does your organisation intend to provide access to digital information into the future?

Digital Preservation Costs

48. Did your organisation attempt to undertake a cost benefit analysis concerning its investments in preservation?

49. Has this analysis been assessed in light of your actual preservation activities? Did it prove to be accurate?

50. To which section of the budget are the economic resources for your preservation programme allocated?

51. What percentage of the organisation's budget is spent on preservation? Can you compare that to some other area of the organisation's activity?

52. Is the organisation attempting to address amortisation issues in the preservation budget?

53. Are there available sources of funding within the _____ sector allocated for digital preservation issues?
☐ Yes
☐ No
If yes, please specify.
____
____
____
____
____
____

54. Are you satisfied with these cross-sector services?

55. If no, what would you like to see available (i.e. what would you think could best be solved in common in your sector)? Would you be willing to engage financially in such information?

56. Are there other external sources available for digital preservation activities (e.g. government grants, cross-sector funds)?
☐ Yes
☐ No
If yes, please specify.
____
____
____
____
____
____

Monitoring of Actions

After having identified what has been undertaken in your organisation with regard to preservation activities, we would like to find out about how these efforts have been monitored.

57. Is the preservation process audited on a regular basis?

58. Is compliance to policies, standards, and strategies audited on a regular basis?

59. Is compliance to other requirements (legal, business etc.) audited on a regular basis?

60. How often are checks made to the preserved material (e.g. for signs of deterioration)?

61. Please specify the criteria used for these audits.

62. Who performs these audits (e.g. internal/external)?

Future Requirements

We would like to ask about the areas in which there is a need for additional attention in your organisation and the sector as a whole.

63. How long do you predict that your current preservation policies, strategies, and solutions will meet your organisation's preservation needs?

64. Is the amount of money allocated for preservation going to change in the future? Will it need to be changed?

65. If more funds were available, what could/would they be used for?

66. What conclusions has your organisation come to about its preservation efforts? Are these satisfactory?

67. What preservation efforts are remaining to be addressed within your organisation?
☐ Further data to be preserved
☐ Revision and adjustment of preservation policies and strategies
☐ Additional resources dedicated to preservation
☐ Technological solutions
☐ Other (please specify)

68. Would you like to see more cross-sectoral or intra-sectoral activity with regard to preservation?

69. Are there any other areas in which you would like to have more information made available on digital information? Where do you expect this information to come from?

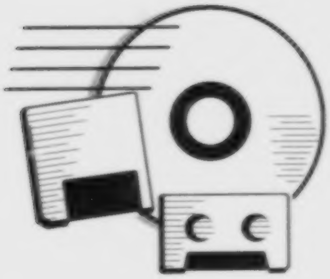
Thank you very much for your valuable contribution.

Résumé

L'Electronic Resource Preservation and Access Network (ERPANET) est une activité de préservation essentielle du 5e Programme-cadre de la Commission européenne. Au nombre de ses autres initiatives, le Réseau a mené récemment une série d'études de cas visant à améliorer sa connaissance des pratiques de préservation de documents numérisés dans les institutions publiques et le secteur privé en Europe. Les données qui ont servi à l'étude ont été recueillies entre les mois d'août et novembre 2002 dans le cadre d'entrevues auprès du personnel de 20 entreprises dans les secteurs de la pharmacie, de la radiodiffusion, de l'édition et aes télécommuniations. Cette communication décrit les méthodes employées pour effectuer cette recherche et en présente les résultats; elle précise également les domaines (recherche, lignes directrices, pratiques améliorées) où des efforts supplémentaires doivent être faits d'urgence.

PRESERVATION STRATEGIES FOR ELECTRONIC RECORDS

STRATÉGIE DE PRÉSERVATION DES DOCUMENTS ÉLECTRONIQUES



Implementation of the Victorian Electronic Records Strategy (VERS) within the Government of the State of Victoria, Australia

Howard Quenault

Public Record Office Victoria
Level 2, Casselden Place
2 Lonsdale Street
Melbourne 3000, Victoria
Australia
Tel.: +61 3 9285 7965
E-mail: howard.quenault@dpc.vic.gov.au

Abstract

The Victorian Electronic Records Strategy (VERS) has evolved as a practicable, effective solution to the problems posed by electronic records in the information age. With VERS, the State of Victoria in Australia has taken a proactive and innovative approach to a global problem. From the original VERS project in 1997–1998, through a pilot implementation in the Department of Infrastructure (DOI), and now with a roll-out across the Government of Victoria, VERS has grown to be a world-leading electronic record-keeping solution with its eyes on tomorrow. The implementation within DOI has allowed the VERS standard and approach to be tested in a large departmental context. The lessons learnt through this process are being used to assist in the continuing roll-out across, initially, another nine departments. Supporting this process, a VERS Centre of Excellence has been established to provide both continuing leadership, development of markets, and the establishment of a long-term digital records archive.

Early initiatives

The issues related to poor management of electronic records are now starting to become visible as a growing number of high profile "horror" stories hit the press. For example:

National Aeronautics and Space Administration (NASA) data from 1970s lost due to "forgotten" file format — In 1999, University of Southern California neurobiologist Joseph Miller asked NASA to check some old data the Viking probes had sent back from Mars in the mid 1970s. Miller wanted to find out whether or not certain information on gas released by Martian soil, at the time dismissed as meaningless "chemical activity," was actually evidence of microbial life. NASA found the tapes he requested, but they didn't find any way to read them. The data, despite being only about 25 years old, were in a format NASA had long since forgotten. Or, as Miller puts it, "The programmers who knew it had died."

Premier's Correspondence Unit (PCU) of the Government of Ontario — The PCU went "paperless" in 1993 using Megadocs software from Digital. At the initiation of this change, it was decided that the archival record would be created and managed by the PCU on behalf of the Archives of Ontario. Three years later, when Digital announced that it would no longer support Megadocs and recommended alternative software, a variety of problems surfaced during preparations for the transition: there had been no archiving of records during the period and all records were online; the migration to the new software would be extremely expensive; the PCU had no hardware platforms to support the Megadocs records and Megadocs itself was now obsolete. As a result of these problems, the Archives of Ontario faced the complete loss of 3 years of a key record series.

We don't need to reflect long on these stories, and the plethora like them, to realize how critical the issue of long-term digital preservation is becoming. In the public context, keeping public records for the long term is vital to the business of government. Probity and good governance demand that governments be able to prove their actions and the contexts in which those actions took place. The day-to-day business of government requires access to the memory of the polity in order to be truly effective. In the longer view, too, public records are an irreplaceable part of the heritage of any community and its citizens.

Recognizing the problem of keeping digital records and archives for the medium to long term, the Public Record Office Victoria (PROV) initiated VERS in 1997–1998. Three key factors were seen as the main drivers for seeking an innovative solution to the problems associated with digital records:

- the rapid pace of technological changes that cause the equally rapid technological obsolescence of computer systems (hardware, software, and media) and the potential loss of the important information within these systems
- the lack of a recognized global strategy for effectively dealing with electronic system redundancy (the scale of the problem, daunting even in the 1980s, has continued to grow exponentially; in the absence of a pre-existing solution, the Government of Victoria had the choice of being proactive in finding a solution or risking the loss of countless bytes of invaluable data)
- the long-term commitment of the Government of Victoria to providing government information and services electronically to all Victorians (thus the likelihood that an increasing volume of government business would be conducted electronically, increasing the scale of the problem)

The VERS project was also a response to the findings of an earlier report, *Keeping Electronic Records Forever* (1996). In that investigation, it was found that there was no adequate worldwide solution to the problem of keeping electronic records in an accessible, contextualized format for the long term. Until then, most recommendations had centred on the migration of data to new applications. PROV had the challenge of finding an application-independent solution that could create "snapshot" records, complete with their context — i.e. electronic records that would be capable of being read in the long term.

The initial VERS project team completed its work in 1999 and published a final report, which included recommendations on the structure of a long-term electronic records preservation strategy for the state. As part of the project, the team developed and demonstrated a prototype electronic records and archives model. The model has been recognized both nationally and internationally as a leading-edge solution to an international problem.

What is VERS?

VERS is a framework of standards, guidance, and implementation projects centred around the goal of reliably and authentically archiving electronic records created or managed by the Government of Victoria.

The VERS standard is a three-part solution for preservation of electronic documents:

- *Long-term format* — Electronic documents (generated by software such as Microsoft Word) are converted to a long-term format (pdf or tiff). These formats were chosen because it is expected that in the future they will provide the best chance that an attempt to read the document will be successful (i.e. a person will be able to view the document in its original form). Unlike Microsoft Word, which is a proprietary format, pdf and tiff formats are in the public domain and are therefore likely to remain readily available to read files.
- *Encapsulation* — One or more documents are "bundled" together with a detailed set of metadata which hold information about the documents (such as when, where, how, and by whom the document or documents were created). The metadata are to be stored using xml (Extensible Mark-up Language) format.
- *Digital signature* — A public-key digital signature is attached to the "bundle" of documents, to ensure that the integrity and authenticity of each electronic document is preserved. The product of this process is known as a "VERS Encapsulated Object" (VEO).

The VERS standard contains detailed technical information about the type of system and functionality required to be able to create and manage VEOs, the metadata fields required, and the exact specifications for how the xml, pdf, and digital signatures should be fitted together to form a VEO.

Guidance on the strategy is supported by an extensive training program, the VERS Web site (www.prov.vic.gov.au/vers), the VERS Web toolkit, agency advice, and consultancy assignments.

A number of implementation projects have been undertaken to provide practical experience in adopting the VERS approach:

- Keeping Electronic Records Forever (1996) addressed the issue of how to ensure the retention of, and ongoing access to, digital records created today and in the future
- Demonstrator System (1998) demonstrated that it was practical to capture, manage, and discover digital records using a VEO format
- the Final Report of VERS (1999) published the findings of the initial VERS project
- VERS@DOI Pilot (1999–2002) implemented the VERS strategy within DOI (this included record capture at creation, management of the VEOs in a VERS-compliant record-keeping system, and the ability to export these to PROV when required)

- the VERS Centre of Excellence (2002–2004) provided further support for the VERS program to roll-out across the nine remaining major departments of the Government of Victoria

Implementation of the VERS solution at the Department of Infrastructure (DOI)

In 1999, DOI was selected to implement a VERS-compliant record-keeping solution as a pilot project for the Government of Victoria. DOI is now in the process of utilizing the capability across the department to capture electronic records and produce an electronic records repository.

DOI has made the intellectual property created in the course of this project, including some software, available to other agencies as they embark on their implementation of the VERS solution. As part of the adoption of the VERS standard across the rest of the Government of Victoria, DOI will provide ongoing advice and mentoring to other departments, streamlining processes and bringing the entire government into a VERS-compliant network.

Lessons learnt

Having a standard format for an electronic record makes sense

The existence of the VERS standard long-term format for electronic records enabled the department to focus its attention on the capture, management, and access of electronic records rather than on defining a standard.

However, even with the standard already defined, there was a significant amount of work necessary in terms of understanding how to apply the standard in a practical way. Therefore, throughout the implementation, compromises had to be made to the theoretical concepts embodied in the standard.

One example of a compromise was in the implementation of digital signatures applied to the record. The standard called for government officers to sign the record with their own personal digital signature, for non-repudiation purposes, as well as to sign it at a departmental level. However, at the time of implementation, the Government of Victoria had not implemented a full Public Key Infrastructure (PKI), making individual signing impractical. Therefore, it was decided that records should be signed on behalf of the department but not individually, the outcome being that the digital signature can be used to prove the integrity of the record but not for non-repudiation purposes.

Implementing leading-edge solutions often brings project complications

One specific goal of the project was to utilize as much commercial-off-the-shelf (COTS) software as possible, thereby reducing the risk of developing an "orphan" solution. Unfortunately, at the beginning of the implementation process, it was found that the commercial marketplace had yet to integrate the VERS requirements into their base products, particularly in the areas of xml, metadata, and digital signatures. This situation improved over the period of implementation with the information technology (IT) marketplace gradually embracing these concepts — leading to an increase in the number of possible software solutions. However, achieving the COTS goal has meant extending project deadlines (to accommodate the commercial products development cycle) and developing additional bespoke integration code.

Awareness and getting your mind around the issue of electronic records

The DOI experience suggests that there is often little awareness of the issues related to electronic records. As business procedures were moved from paper-based to electronic processes, the business community generally expected that the software would automatically solve their records problems. In reality, however, the business records that were generated were not, in most cases, stored or managed in a way that provided the necessary accessibility.

As business applications are being replaced more and more frequently, the need to guard records against system obsolescence is becoming more and more urgent. In most environments, however, the fact that the problem will surface "some time in the future" means that resources are often not applied to mitigate the risk now.

These issues can be addressed in a number of ways:

- by raising awareness of the issues
- by designing automated electronic records capture into new business applications to make it easy for users
- by identifying business opportunities to start the process of capturing electronic records, i.e. at major milestones within projects
- by determining and implementing the simplest approach to the capture of the electronic record

Open data

Many software applications base their longer-term survival on the ability to manage application data within their own propriety data formats. This strategy requires customers to migrate data through continual software upgrades. But adopting the VERS standard format allows the capture of the record as close to the point of creation as possible. It also allows the data to be kept in an open data format for a significant amount of time without requiring continual format changes.

During the course of the project, the decision was made that, as a means of supporting re-use and accessibility for the long term, both the original data format and a long-term format should be held for each digital object. Use of an open data format has meant that the record does not rely on a specific application, software, or hardware to retain its accessibility and value.

Knowledge exchange
To help raise the level of awareness both within DOI and across the rest of the Government of Victoria, a community of practice was established. Comprising records managers, information managers, knowledge managers, IT professionals, and business representatives, the group exchanged information on a large number of issues and devoted a significant amount of time to the distribution of information: via the VERS Web site, in face-to-face discussions, and through presentations and industry briefings.

Adoption of VERS in an e-business context
During the last few years, there has been a significant loss of the record-keeping ethos across government. However, with the movement to e-business, it will be important to consider and incorporate records requirements (for both hard-copy and electronic records) while the new Web-assisted business processes are developed.

At DOI we found that, during the push to get business online, there was often significant pressure to deal with only the immediate project problems — and the requirement to capture, manage, and access electronic records tended to get left for later. In this scenario, there is a very great danger that a significant number of electronic records will continue to be generated without regard to their management, retention, and accessibility. The outcome of this situation is that many records will ultimately be lost forever.

Recognition of cultural change
Most government workers, like most businesspeople, want to get on with the "business of doing business" and expect the IT infrastructure to capture records automatically. While this may be true in the future, we are currently in a liminal period between the predominantly paper-based record era and the primarily "born-electronic" era. During this transition it is common for old practices of paper record-keeping to be lost before good electronic information management principles become part of the cultural norm. Recognizing this organizational reality means that pragmatic steps are required to better manage electronic records. Even such apparently mundane devices as introducing file-naming conventions can have a substantial impact in better information management.

Initiatives that were undertaken in the VERS@DOI project included restructuring disk directories to reflect a functional business classification scheme, creating an identifying link from paper files to electronic records, and standardizing file-naming.

Impact of technology changes in the environment
Implementing an electronic record-keeping environment is not a once-off project. Capturing records within an e-business and workflow framework requires integration with desktop tools, back-office applications, and specialized technology. During the VERS@DOI project, DOI migrated:
- from Microsoft Windows NT to Windows 2000
- from SQL V7 to SQL 2000
- from Microsoft Office 97 to Microsoft Office XP

and also implemented:
- a number of Lotus Notes workflow systems
- a number of document management systems
- a sophisticated Web environment covering Internet, extranet, and intranet

These major changes in the IT infrastructure had a significant impact on the ability of the business units and the project team to adapt to the continually increasing requirements. It was therefore decided to develop a simple graphical user interface (GUI) that could capture any type of file format from the disk drive, supporting the lowest common denominator to be used in solution construction.

It also became clear during the implementation process that an application programming interface (API) was necessary to ensure we could successfully integrate the record-keeping system with business workflow centric applications.

Implementation approach
After embarking on a roll-out and training program for a substantial pilot project within DOI, it was discovered that the storage of records was minimal.

The primary reasons for this were:

- the view that IT infrastructure would manage records without any user intervention
- the fact that the majority of users didn't see records management as part of their responsibility
- the belief that it was someone else's job / problem

Recognizing this, it was decided to change tack and concentrate on three streams of activity:

- develop a document and records strategy for each division that would engage senior management, thereby obtaining their support to drive organizational change
- identify key business issues with electronic records and develop solutions specifically for these problems
- identify new e-business applications and work with the designers and developers to embed the capturing of records alongside the business process

Electronic record identification and business analysis

The question of what constitutes a record came up time and time again, which again illustrated the changing ethos related to record-keeping. In the course of business activity, a document might go through a number of drafts and versions, and have a number of associated e-mails, briefings, and reports. But which of these electronic objects must be captured and at what point in the business cycle?

It became obvious that the steps in the e-business processes needed to be analysed in detail to determine the electronic record components that needed to be captured and at what point this needed to happen.

Technology – Will it solve the problem?

There has been considerable debate worldwide about whether or not technology will ultimately solve the electronic records problem if we wait long enough. But based on our experiences since 1996, the problem has continued to get worse rather than better. The number of digital objects, the number of repositories, and the number of formats have all continued to increase. Given the exploding volumes we are now dealing with, the availability of a search engine to find information across these repositories and the ability to sort the wheat from the chaff is questionable.

Emerging benefits

At DOI, capturing electronic records in a VERS format has allowed the storage of a growing number of self-described "sealed" electronic records in a consistent format. In the event that these records need to be migrated at some time in the future, there will be significantly less work than if the documents had only been captured in their original format.

Having all records in a VERS format also allows them to be shared within government or across jurisdictions in a manner in which they can be understood, decoded, and loaded into other record-keeping systems. The department can also use the records in business processes where the authenticity and integrity of the record is crucial.

VERS Centre of Excellence

VERS has been mandated by the Government of Victoria as a key plank of the government suite of initiatives, and AUS$8.2 million in funding has been committed to the program over the coming 2 years to see VERS rolled-out across the central government space. A VERS Centre of Excellence was established with this funding in July 2002, and an electronic records repository to hold the permanent electronic archives of the state is due to be completed by December 2004.

The VERS Centre of Excellence has participated in the promulgation of VERS on a number of levels in its first year of existence. Key achievements include:

- substantial research into service delivery models for VERS
- assistance to the vendor community for the adoption of the VERS format with two products becoming VERS-compliant
- development of an intellectual property framework to cover material developed as part of the VERS program
- results from the first VERS-readiness assessment undertaken across the nine inner-budget departments of the Government of Victoria
- establishment of a VERS consultancy panel to provide expert assistance on the application of the VERS strategy
- two further large government departments adopting VERS-compliance, using the impetus of new Electronic Document Management (EDM) system roll-outs to get VERS built in
- an extensive, multi-faceted VERS training program which has educated more than 50% of all Government of Victoria records, archives, and IT staff in electronic record-keeping
- revision and reissue of the VERS standard with greatly enhanced clarity and flexibility
- completion of the VERS digital archive design specification

Future work of the VERS Centre of Excellence will focus on:

- development and implementation of the VERS digital archive
- continuation of the program particularly in relation to support of other Government of Victoria departments as they become VERS-compliant

Beyond Victoria

The VERS Centre of Excellence is and will continue to be a substantial player in world standardization attempts in this area, with representation on Adobe's PDF-A committee and the Records Management committee of ISO. The VERS model itself has been adopted outside the Victorian jurisdiction, with endorsement from Collaborative Electronic Notebook Systems Association (CENSA) and on-the-ground use by a number of multinational organizations.

Conclusion

Over the last 6 years VERS has moved from an idea to a vision and now to a practical implementation.

It has become apparent that there is a need to concentrate on records rather than systems when considering electronic records.

Attention needs to be on the "comprehensive" record of business, including physical and "born-electronic" components.

There is now sufficient practical understanding to support the notion that management of electronic records using a long-term format is a good idea.

To capture, manage, and provide access to the electronic records of business there is a need to retain implementation flexibility and consider innovative solutions like VERS.

Bibliography

Department of Infrastructure. *VERS RKS Strategy – Pilot System*, 2001. www.prov.vic.gov.au/vers/published/ publcns.htm#briefings

Department of Infrastructure. *VERS RKS Analysis & Design Report – Pilot System*, 2001. www.prov.vic.gov.au/vers/published/publcns.htm#briefings

Public Record Office Victoria. *Keeping Electronic Records Forever*, 1996. www.prov.vic.gov.au/vers/published/publcns.htm#briefings

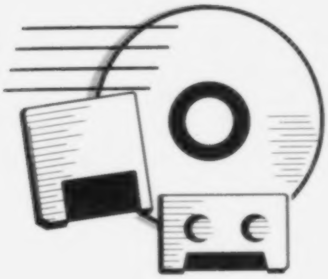
Public Record Office Victoria. *Victorian Electronic Records Strategy Final Report*, 1999. 138 pp. www.prov.vic.gov.au/vers/published/publcns.htm#briefings

Public Record Office Victoria. *Standard for the Management of Electronic Records*, V 1.2 and V 2.0, 2000–2003. www.prov.vic.gov.au/vers/published/publcns.htm#briefings

Résumé

La Victorian Electronic Records Strategy (VERS) est devenue une solution efficace et pratique aux problèmes posés par les documents électroniques à l'ère de l'information. La stratégie VERS a permis à l'État australien de Victoria d'adopter une approche proactive et novatrice à un problème d'envergure mondiale. Depuis le projet initial, lancé en 1997-1998, qui a été suivi par la mise en œuvre pilote au ministère de l'Infrastructure de l'Australie puis par un déploiement dans l'ensemble du gouvernement de l'État, la stratégie VERS est devenue une solution de préservation des documents électroniques de classe mondiale qui est tournée vers l'avenir.

La mise en œuvre de la stratégie VERS à ce Ministère a permis de valider la norme et l'approche dans le contexte d'un grand ministère. Les leçons apprises durant l'exercice sont mises à contribution dans le cadre du déploiement qui se poursuit, au départ, dans neuf autres ministères. À l'appui de l'exercice, un centre d'excellence de la stratégie VERS a été mis sur pied pour offrir un leadership permanent, assurer le développement de marchés et établir des archives numériques à long terme.



VOLATILE MEMORY

ULRICH LANG

Head of Conservation
Museum für Moderne Kunst, MMK
(Museum of Modern Art)
Domstrasse 10
60311 Frankfurt a.M
Germany
Tel.: +49-69-212-33719
E-mail: u.lang@conserve-art.de

Abstract

This paper looks at the issues concerning conservation and exhibition of works of art that include technical elements. Is it possible to differentiate between the content of an artwork, its data, and its manner of portrayal? These questions help to determine what elements are fundamental to the artwork — information that is especially important when dealing with electronic media art whose parameters are complex and difficult to define but also essential for the accurate exhibition of the work. These issues are discussed in relation to Nam June Paik's closed-circuit installation *One Candle*.

"I love anti-technological technology"
(Nam June Paik 1988).

Many artists express themselves in media that extend far beyond paintings and classical sculpture. In fact, there is really no material or method of expression that cannot be part of an artwork or installation, and each medium brings with it a variety of parameters for preservation and conservation. This begs the question: What makes an artwork authentic "when its meaning has shifted away from that of an unique and precious object?" (Laurenson 2001). Does installation art by its very nature deny the possibilities of preserving it? In the 1960s, groupings such as *Fluxus* and *Eat Art* ran contrary to the established art scene of the time. This resulted in very few of these artworks being recorded and collected — and even if they were, it was just the outcome or residues that were preserved. Indeed, today almost any of their artifacts (e.g. posters, tickets, etc.) are kept and become precious. Although these items were never intended to be art objects, in many cases they are all that remains of the works.

Nowadays an exhibition or museum is often the very platform on which an artwork is created. In many cases artists start with an incomplete plan that evolves and shifts as they work within the site (Real 2001). But is there a certain moment at which the opus is "finished" and handed over to the public and its fate even though it seems to be arbitrary. When radical German painter Günter Umberg was asked "When is your painting ready?", he answered "When I am close to it" (Umberg 2002).

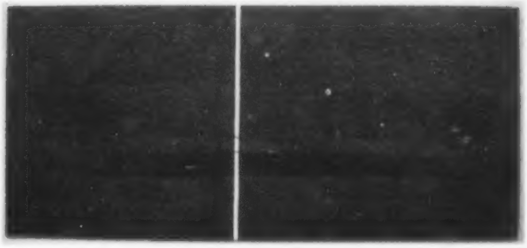
Artists often leave it to others to determine which aspects of the artwork are essential for showing it. But what *is* fundamental to a work's portrayal? When talking about artworks that include electronic media, this means first defining all the parameters that are necessary to show the work of art. For example, is a video or program code stored on the record the only information or does it also include the equipment that is necessary to produce the video or data stream? And does the artist dictate a certain size and appearance of the monitor or indicate if a visible picture should be projected to a wall, screen, or other surface? What should be done if an original black-and-white monitor is no longer functioning? Can it be replaced by a flat screen which undoubtedly produces a different picture?

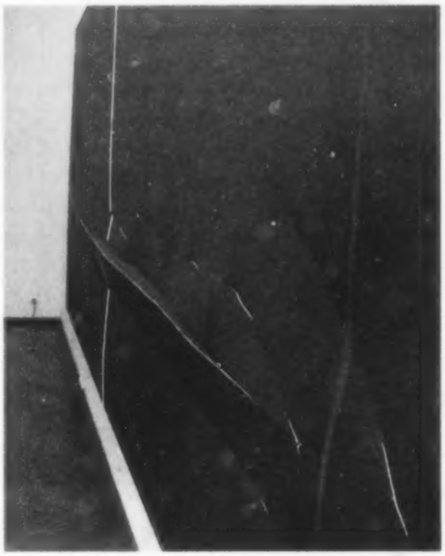
The IIC-CG *Code of Ethics* proposes that no aspect of a cultural property should be altered nor should material be removed from it without justification. This runs contrary to the common practice of exchanging or borrowing parts of technical, electric, and electronic media. It goes without saying that equipment such as plugs, monitors, slide projectors, disc players, etc. is interchanged, modified, or replaced as necessary. Although these parts seem to be exchangeable, for coming generations they will indisputably be the authentic victims of the artwork. No matter what medium artists use, it is always a challenge to preserve it — a challenge that may be time-consuming and expensive, and one that is often compounded by the lack of a clear preservation concept. Of course everything should be kept in the best possible condition, but it is also necessary that a certain amount of time-based patina and degradation be accepted, even for technical installations or devices.

A lot of articles and books provide good advice about storing and using electronic records, and there is currently a lot of investigation being carried out about the best ways to restore tapes and discs. But how do we detect defects in an installation? Is it a fault if the environment is brighter or darker, if the light temperature is different, or if the sound level is not as the artist intended? At what point should we intervene in the case of fading monitors or projectors? In most cases the acceptable parameters for decay remain undefined. Perhaps we are too used to the availability of electronic media, which seem to be cheap and easy to handle with very few worries about packing, insurance, storage costs, etc.

A lot of time and money is often spent on restoring a painting (see Figures 1 and 2), which might take days or even months. To prevent unnecessary damage, it is well accepted that light-sensitive organic material can be exhibited safely for only a limited amount of light or time (Laurenson 2001). But it is less well known that video installations (because they are subjected to friction while being played) should also be displayed on a limited basis. And if damage should occur, there seems to be no doubt about the validity of retouching a painting. But what about re-phosphoring a TV tube? Might it be an option to transfer the magnetic layer of a tape or the metallic layer of a disc to save it?

When any part of a video installation is changed — whether it is the record itself, any part producing the video stream, or the environment — it is possible that the appearance of the artwork might be altered dramatically. But is it possible to make a distinction between the *content* of the record and the *portrayal* of an artwork, between the data and the transmitter (player, computer, projector, etc.)? Just as the support for a painting is an important part of its structure whether or not it is visible, the audiovisual playback set of an artwork that comprises electronic media is part of its structure. The expression "video" in general means the data stream, and the visualization of information that can only be read with technical equipment. Video today is a high-frequency and therefore dense data stream in analog or digital form. But the fact that an artwork is composed of technical elements does not automatically justify technical progress. What happens to a video that is digitized (and in most cases compressed) if the next generation of the same compressing software uses slightly different parameters? A colour shift would be the simplest change. As transistors and processors become faster, so does the image. That becomes evident if one transfers an early computer-animated video or video game to a modern computer — not to mention the difficulties with the platform and operation that occur with such a transfer. And every transformation can cause a chain reaction for the equipment and lead to a completely different final appearance. This makes clear the need for documentation not only for the record but also for the necessary equipment to visualize the content of the record — including all technical details such as connection scheme, circuit diagram, plugs, sockets, etc. What art historian Cesare Brandi formulated in 1963 in his *Teoria del Restauro*, i.e. "the principle that the original material identity of an object should





Figures 1 and 2. Barnett Newman's Cathedra *(top), which was cut to ribbons on 21 November 1997 at the Stedelijk Museum (bottom). It took 4 years and a group of some 30 researchers, advisors, conservators, and technicians to restore the painting before it could be given back to the public on 8 December 2001.*

always remain the most important source for ever-changing interpretations" (van Wegen 1999), is therefore still true.

Although restorations of classical paintings or even modern materials such as polyurethane foam are undertaken on a very high technical and scientific level, conservators usually leave the electronic media and the record itself to technicians. And discussions about the equipment and technique (e.g. what monitor or is projection allowed, etc.) have not even started yet. Small repairs are often done without being documented, and unfortunately the equipment of one installation is often used for others.

Therefore, if we decide that an artwork should not just become another version or a later edition, it is necessary to confront the question of when to intervene. Without proper documentation eventually nobody will be able to remember what the authentic system was, or if it is known, the original type of system will no longer be available on the market. The question must therefore be addressed as to whether or not it is best to spend a lot of time and money to find or repair original devices or just buy new equipment. But video installations can also come to grief because of simple questions like the size of the original monitor, installation plans, etc. One of the first things done when documenting a painting is taking its dimensions, but who is measuring similar details for video installations? For example: How big is a picture projected onto a wall? How does a black-and-white image look on a colour TV? What is the quality of the sound? And is a break that occurs in a tape while being rewound part of the artwork or at least part of the feeling?

On the other hand, does a video installation work at all if the video does not work? Is it therefore legitimate or even a necessity to keep it going at all costs, or are there limits set by the properties of the materials used? Is it allowable or mandatory to exchange parts due to wear and tear? If this is done, where does retouching end and replacement start (Lang 2002)? Thus it is necessary to define the parameters in which compromises are acceptable or what the minimum level should be in order to document and preserve a work of art in its integrity. That brings us to the point where we have to think about the electronic record as such. Is it enough to transform the program code, sound, or video signal to the latest standard not to mention the discussion about the longevity of a record, or might it be necessary to keep the original version for display?

Some of the problems in conserving video art can be demonstrated by looking at an installation by Nam June Paik (probably the first video artist). When the Museum for Modern Art (MMK) opened in Frankfurt in 1991, a number of artists were invited to install works in the museum. One of these artworks is Paik's closed-circuit installation called *One Candle*. There are — from a technical point of view — other examples in the collection that are more complex, but this simple work is helpful in discussing the subject on a basic level. The installation has two precursors: the *TV Buddha* from 1974 (which shows a Buddha filmed by a camera while watching himself in the TV) and the *Candle TV* from 1975 (which is a burning candle in the hollow shell of a 1950s TV set, which looks like an antique relic from some bygone high-tech age).

Paik's *One Candle* consists of the flame of a candle that is constantly filmed by a video camera and projected onto the walls of a small triangular room by six projectors (Figure 3). The piece was first shown in the Portikus, an exhibition hall in Frankfurt, with slightly different equipment.

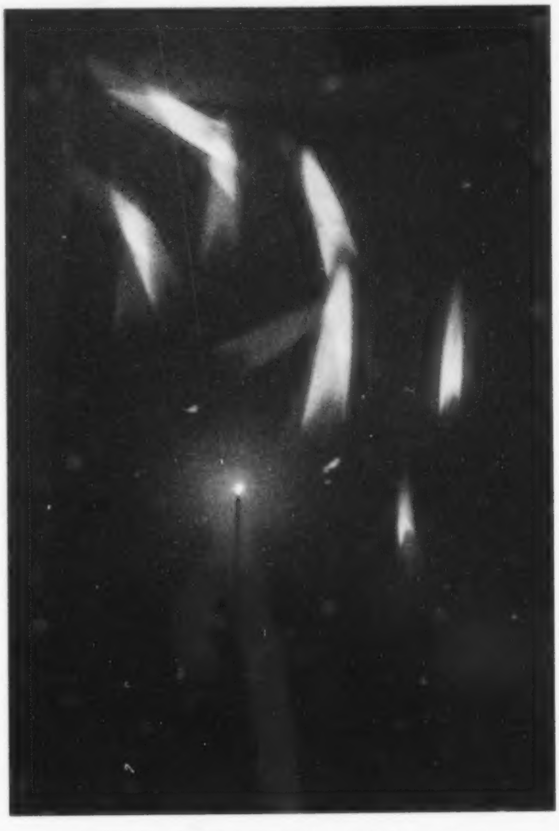


Figure 3. Paik's One Candle *installation from 1989 to 1996, before the projectors were exchanged.*

Paik gave instructions for the MMK version and his assistant, Jochen Saueracker, installed it together with the museum staff. Paik first saw the installation at the MMK, when it was already set up.

The installation at the MMK in 1991 originally existed of the following: the burning candle on a speaker tripod; a video camera on a camera tripod; one control monitor; a video distribution amplifier; one transformer (220 V to 110 V; 1000 W); six cathode ray projectors of two different models (two by Advent and four by General Electric); and various cables and plugs. The "information" does not need to be stored on any record but is the burning candle itself (replaced every day). The idea of the artwork is the candlelight and its reflection or better projection on the wall. According to Jochen Saueracker, it is important for the installation to have a central wall with two projections above each other that overlap and create a white area in the middle (Figure 4). Candle and camera should be a central element and visible. The projectors are within the room and therefore have a sculptural presence. For his installations Paik often uses low-end equipment (Saueracker 1991, 2002).

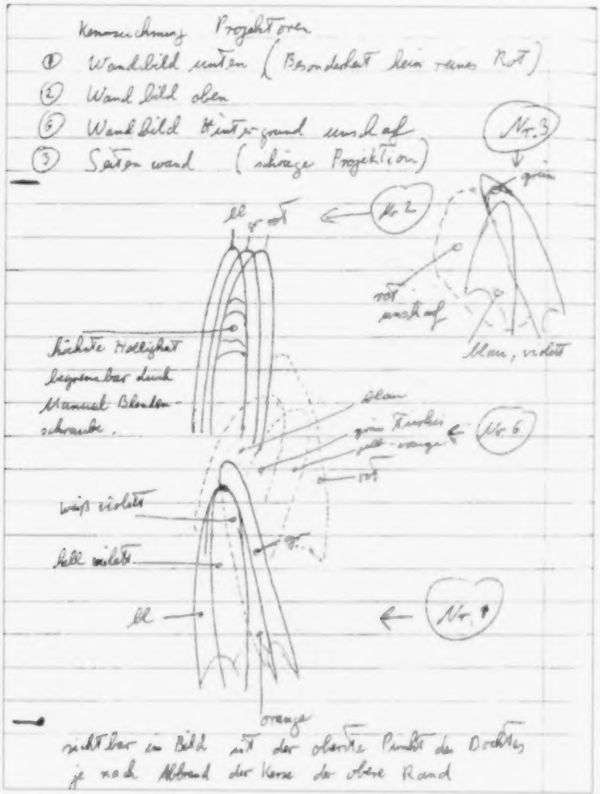

Kennzeichnung Projektoren
① Wandbild unten (Besonderheit kein reines Rot)
② Wand bild oben
⑥ Wandbild Hintergrund unscharf
③ Seitenwand (schräge Projektion)
Nr. 3
grün
Nr 2
bl
gr
rot
rot unscharf
blau, violett
höchste Helligkeit regulierbar durch Manuel Blenden-schraube.
blau
grün Türkis
gelb-orange
Nr 6
rot
weiß violett
gr
hell violett
bl
Nr. 1
orange
sichtbar im Bild ist der oberste Punkt des Dochtes
je nach Abbrand der Kerze der obere Rand

Figure 4. Jochen Saueracker's sketches for the first MMK installation.

The piece has now been on display for more than 12 years, which translates to a total of more than 30 000 hours of actually being shown. It is evident that the original video equipment has not run for so many years without needing repairs. The control monitor, once used to control the position of the candle, is not used any more, the video camera has been exchanged by a spare one Paik supplied, and the upkeep of the projectors has been very time-consuming. At least once a week a technician has had to repair the projectors, tubes have had to be replaced, the circuit boards have had to be soldered, etc. When, in 1996, Nam June Paik was asked how to maintain the installation, he agreed to a new set of projectors. The original projectors were replaced with a newer generation of a very common model, the Sony VPH 1000 Series. The museum was able to pay for two of the necessary projectors ($4000 each) and Sony sponsored the other four. Paik's assistant Jochen Saueracker carried out the replacements. The room, the light, and even the shelves for the projectors remained the same. Even though he used the same positions and stands for the projectors, the image of the work (e.g. colours, sharpness and dynamics, sculptural aspects, etc.) changed completely (Figure 5). Another aspect that changed (and one that is always underestimated) was the sound quality; in this case the original projectors had no humming ventilation system — unlike the replacement projectors.

Paik had signed one of the original projectors, so it was left in the middle of the room and today looks like a leftover from an ancient video era. Because Paik was involved in the changes to *One Candle*, the 1996 installation can probably still be considered authentic. But are there now two different versions at the MMK after the installation in the Portikus? Or is the artwork just an idea that can be reproduced with any equipment — like a multiple? In contrast to Paik, who did not say anything about the dimensions or the room itself, artist Bill Viola gives precise instructions and floor plans for his work (Figures 6 and 7). When showing one of his pieces the room must be modified to match his requirements.

Today we are in a similar situation to that of 1996 with regard to Paik's installation. Some basic repairs are necessary, all cathode ray tubes are dramatically faded, and the normal maintenance costs for the projector are about $10,000 every 2 years. The total of all the recent repairs (some ventilators, a circuit board, some tubes, etc.) indicates that another replacement will soon be necessary. But what kind of projector would that be? The cathode

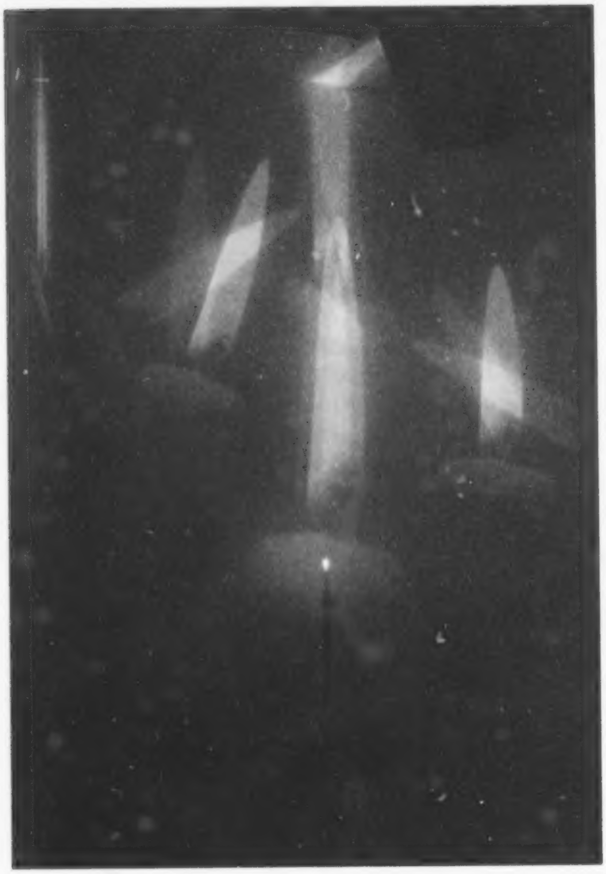


Figure 5. Paik's One Candle *installation since 1996 with the new set of projectors.*

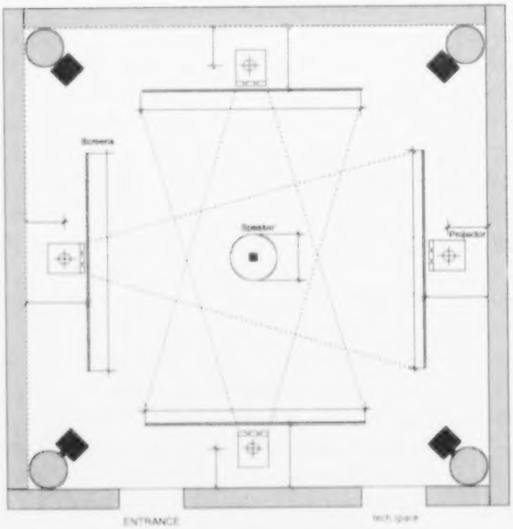


Figure 6. Viola's Stopping Mind *installation plan, Viola Studio.*

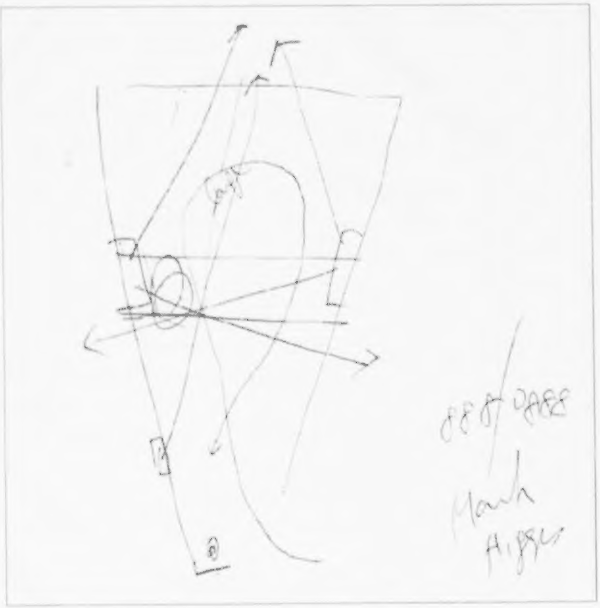


Figure 7. One Candle *installation plan, attributed to Paik.*

ray projectors that Paik originally used are no longer produced (only a few are still available), and an LCD projector with only one tube could not project three different images of a burning candle. Because it is important to have the overlap on the wall (the overlap produces a whole colour range and, as a result of the overlap of the three colours, a white image in the centre), one option would be to put three small LCD projectors in one old body, one each for red, green, and blue. But this solution would have enormous costs, would destroy the interior of the old projectors, and would reduce the possibilities of future reconstruction. And the image would still be different, even if filters were used to match the light temperature.

We know that Paik has nothing against replacing parts but he says: "Everybody can make such a sculpture, but I sign it. When I am dead, it is your problem to find out which is the original. I have two originals: one piece and one copy of a better quality" (Paik 1995).

But who in the future will be in charge? Who will take the responsibility? Paik leaves it to us to find out what is the original. Today, as we are still close to the technique, it seems to be no problem to change the equipment because we can recall the feeling of the original or authentic set. But the next generations want to recall the equipment and materials used in the late 20th century. Asking the artist is one aspect, but we should not hide behind that. We should take over the responsibility even if it might be criticized. Hans Ulrich Reck, a contemporary German media-philosopher says (Reck 1995):

> Inasmuch as 'guaranteed', has to be understood under 'authentic', the conservator and the restorer become decisive agents and instances of the authentic. They are not only guarantors of that which their work aims to conserve, but also actors who join in the construction of whatever this guarantee embodies. The guarantee must be reliable, creditable, granting insight into the relation to the original, it must secure attributes.

Because the equipment from Paik's original installation in 1991 was still in the warehouse of a shipping company, the museum was able to acquire the original projectors. We will now try to repair or restore them instead of emulating a new set of projectors. Even though they were replaced in 1996 by the set that is on display until today (and therefore longer than the first set), and even if they may need more maintenance than new projectors, their appearance might convey a more authentic feel for what the installation looked like in 1991. The expenditure to keep and repair old technology or to emulate and modulate it is probably the same anyway. But if we have learned anything in dealing with technology-based media it is that *all* elements (information, systems of transmission, and physical wear) are strictly linked within an ever decreasing circle of time. All aspects are intimately connected both technically *and* aesthetically, and any modulation will have complex consequences.

Bibliography

Lang, U. "Motion as Material — The Kinetic Art of Gerhard von Graevenitz." pp. 165–176 in *Plastics in Art, History, Technology, Preservation* (edited by T. van Oosten, Y. Shasoua, and F. Waentig). München: Kölner Beiträge zur Restaurierung und Konservierung von Kunst- und Kulturgut, Band 15, 2002.

Laurenson, P. "Developing Strategies for the Conservation of Installations incorporating Time-based Media with reference to Gary Hill's 'Between Cinema and a Hard Place'." *Journal of the American Institute for Conservation* 40, 3 (Fall/Winter 2001), p. 263.

Paik, N.J. "Interview with the Artist." p. 107 in *How Durable is Video Art? Contributions to Preservation and Restoration of the Audiovisual Works of Art.* Symposium at the Kunstmuseum Wolfsburg, 25 November 1995.

Real, W.A. "Towards Guidelines for Practice in the Preservation and Documentation of Technology-based Installation Art." *Journal of the American Institute for Conservation* 40, 3 (Fall/Winter 2001), p. 212.

Reck, H.U. "Authenticity in the Fine Arts (to the Present Day)." p. 85 in *How Durable is Video Art? Contributions to Preservation and Restoration of the Audiovisual Works of Art.* Symposium at the Kunstmuseum Wolfsburg, 25 November 1995.

Saueracker, J. Installation sketch from 1991; interpreted in 2002.

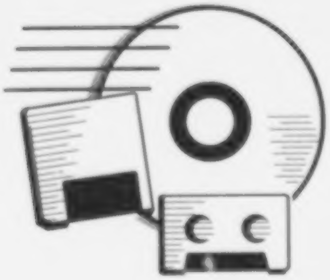
Umberg, G. Personal communication, September 2002.

van Wegen, D.H. "Between Fetish and Score. The Position of the Curator of Contemporary Art." p. 202 in *Modern Art: Who Cares? An Interdisciplinary Research Project and an International Symposium on the Conservation of Modern and Contemporary Art* (edited by I. Hummelen and D. Sillé). Amsterdam, 1999.

Résumé

Nous examinerons les questions relatives à la conservation et à l'exposition d'œuvres d'art comportant des éléments techniques. Est-il possible de faire une distinction entre le contenu d'une œuvre d'art, ses données et la façon dont elle est présentée? Il s'agit de considérer ce qui est indispensable à la présentation de l'œuvre, notamment lorsqu'il est question d'œuvres sur support électronique dont les paramètres sont à la fois complexes et difficiles à cerner, mais nécessaires à l'exposition de l'œuvre. Nous parlerons de l'installation en circuit fermé de Nam June Paik intitulée One Candle *(« une chandelle »).*

I love anti-technological technology
(« J'adore la technologie anti-technologie »)
(Nam June Paik, 1988)



ROBERT CREELEY'S COMPUTER FILES: A COLLABORATIVE APPROACH TO LINK ACCESS WITH PRESERVATION

MICHAEL G. OLSON

Special Collections Assistant Librarian for Electronic Media Materials
Stanford University Libraries, Department of Special Collections
Green Library, 557 Escondido Mall, Room 351H
Stanford CA
USA
Tel.: (650) 723-1704
E-mail: mgolson@stanford.edu

Abstract

The collection of Robert Creeley's Computer Files has occupied a prominent role at Stanford as a test bed project to develop new preservation strategies for born-digital electronic records.[1] From the project's inception, the objective has been to develop and assess the effectiveness of a methodology for incorporating accessibility into digital preservation. The term accessibility is used here in the context of making a digital resource discoverable and providing a means by which interested scholars and the public can interact with this resource. This paper will highlight how preservation and access decisions are inseparably linked and require a greater understanding of how patrons will use electronic resources. The issues of staffing, migration versus emulation, authenticity, intellectual property rights, and privacy will all be discussed in the context of both preserving and providing access to this unique resource.

Introduction

American poet Robert Creeley was an originating member of the "Black Mountain" school of poetry during the 1950s. He was and continues to be an avid and early adopter of the computer as a compositional tool and a medium for correspondence. A collection of Robert Creeley's Computer Files (consisting of fifty-one 3½-inch floppy disks containing electronic working drafts, business documents, software programs, faxes, and correspondence from 1988 through to 1999) was acquired by Stanford University Libraries in 1999. At the beginning of 2002, the Department of Special Collections of Stanford University Libraries embarked on a project to preserve and provide access to this electronic resource. Since the project's conception there has been an urgent need to develop, implement, and assess a real-world solution to Stanford's growing body of born-digital collections.

Project methodology

Staffing

The allocation of sufficiently skilled staff is crucial to the design and development of any digital preservation project. To meet the needs of its own growing electronic resources, Stanford University recently created two new specialist librarian positions. These positions are in different units of the libraries, but have overlapping and cooperative responsibilities for preserving and providing access to electronic resources.

One of the new positions, Media Preservation Librarian, is in the Media Preservation Unit of Stanford University Libraries.[2] This position is staffed by Ms. Hannah Frost, and in this role she is responsible for electronic media preservation for all of Stanford's libraries. She is an important advisor in the proper storage and handling of computer hardware, media, and file formats for Robert Creeley's Computer Files and all other Stanford digital projects.

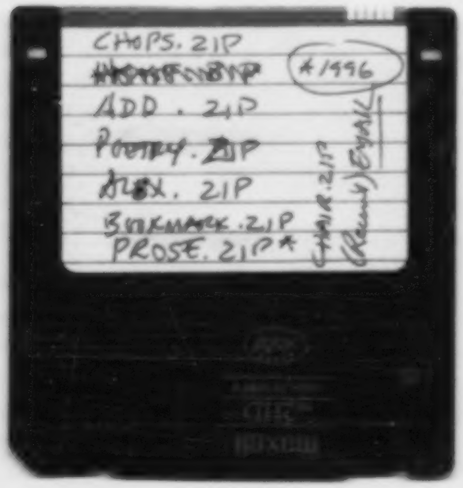
The second position, Special Collections Assistant Librarian for Electronic Media Materials, is in the Department of Special Collections. This is my position, and in this capacity I am responsible for preserving and developing access strategies for all electronic media in the department[3] (the Department of Special Collections contains the largest concentration of unique electronic media materials at Stanford, justifying its need for a professional position devoted exclusively to electronic media).

Although we are located in different units of the Stanford University Libraries, there is extensive collaboration between Ms. Frost and myself.

Robert Creeley's Computer Files is only one of the electronic collections for which a preservation program is actively being developed, and the results of this project will be crucial in forming Stanford's larger digital preservation efforts.

Description of the resource

The collection of Robert Creeley's Computer Files consists of fifty-one 3½-inch floppy disks that are backups of computer files originally created on early IBM-compatible personal computers. Twelve of the floppy disks in the collection are no longer readable, making preservation of this collection a high priority. The remaining readable diskettes contain approximately 12 000 individual files (34 megabytes of data) and span 11 years (1988–1999).



A Robert Creeley 3½-inch floppy disk (Media Id: M1081_S1_SS0_01_027OR). Reproduced courtesy of Robert Creeley and Stanford University Libraries.

Floppy disks: An unreliable storage medium

Floppy disks (or diskettes) are small plastic disks coated with magnetic particles and covered in a protective plastic jacket with a radial slit. They are small in size, portable, and inexpensive. They were an extremely popular storage medium from the 1980s through to the late 1990s, and they are very common in many of Stanford's collections.

But floppy disks are notoriously unreliable as an electronic storage medium. Exposure to strong magnetic fields, drive head misalignments, the composition of the disks, and their physical flexibility all contribute to their high unreliability as a long-term preservation medium.[4]

Refreshing to managed hard disk

The first preservation priority undertaken with Robert Creeley's Computer Files was to refresh the data onto more reliable magnetic media. Stanford hard disk was considered to be the best solution as it is regularly backed up as a safety precaution in the event of hardware failure.[5] As the data on each disk were refreshed, an ASCII (American Standard Code for Information Interchange) text file was created. These files store descriptive and technical metadata for every refreshed Creeley diskette. Each of these files contains the disk label, the disk number, the date that the disk was refreshed, a Media Id, the name of the institution and individual responsible for the transfer, the number of files on the disk, and a notes field. The metadata contained in these files are critically important as they ensure that the descriptive information for each floppy disk is retained along with the technical details of the refreshing process.

The creation of a Media Id is particularly important as this identifier is being used to manage and distinguish between the original floppy disks and the refreshed surrogate data files.[6] The Media Id identifies the digital resource and distinguishes the data it contains as either a unique original or a refreshed copy.

To date, only 39 (or 76%) of the 51 floppy disks have successfully had their data refreshed to Stanford hard disk space. Of the 39 refreshed disks, two required the use of a special data recovery program called GetDataBack by Runtime Software to successfully recover and refresh their files.[7] The remaining 12 floppy disks were sent to a data recovery firm where it was determined that their data were in an Amiga 2000 format.[8] These data will be recovered shortly using an archived Commodore Amiga 2000 from Stanford's Cabrinety collection.

Important lessons were learned while refreshing Robert Creeley's Computer Files. The first is that there is no visual evidence of physical deterioration that can aid in the prioritization of preservation activities for electronic storage media. Visual inspection of floppy disks and many other types of electronic media is problematic and in most cases not very useful in setting preservation priorities.

The second observation is that the label information that accompanied the floppy disks was of little value in determining their contents. A typical example from Robert Creeley's Computer Files is the labeling found on a floppy disk in the collection that identifies it as "Email 2."[9] This minimal label information is common throughout the collection and does not begin to describe the contents, function, importance, or relationship between this floppy disk and the others. Creation dates are often missing and labels are abbreviated, further complicating prioritization. This is particularly noticeable on floppy disks created during the late 1980s through to the middle of the 1990s when DOS operating systems restricted file names to eight or less characters.

Another important factor is the 12 hours of staff time necessary to refresh the data contained on the 39 floppy disks. Copying the data to Stanford hard disk was time-consuming because of the need to scan each disk for viruses and generate the descriptive and technical metadata documenting the refreshing process.

Authenticity

Scholarship in the humanities is based upon interpretation of primary and secondary source materials.[10] This process of interpretation relies upon verification that a record is free from tampering and is what it purports to be. With electronic records the diplomatics or processes for determining whether or not a record is authentic are very different than they are with traditional paper documents.[11]

How do we verify that a body of computer files was created by Robert Creeley? Electronic records are highly malleable and subject to change. The ease with which they can be copied, modified, and manipulated has a profound effect on our ability to verify their authenticity. Which copy is the real copy? Are the copies the same? How does one confirm this information and pass it on to the individual or institution referencing the electronic records? The capture and generation of sufficiently detailed descriptive and technical metadata is a significant step in providing the necessary information to verify an electronic record's authenticity.[12]

Descriptive methodology

Describing the computer files was the next preservation task and a critical component to making the files accessible. When first acquired by Stanford University Libraries, Robert Creeley's Computer Files had no descriptive or technical metadata other than labels. While refreshing the data physically preserves the structure of the floppy disks, it doesn't provide a mechanism for viewing, interpreting, arranging, or describing the files. Almost all of the computer files are in older, now obsolete proprietary file formats that are inaccessible using currently available commercial desktop software.[13]

Emulation

Emulation is a preservation technique that uses modern computer programs to duplicate the functions of old and obsolete computer hardware and software.[14] But visually displaying electronic records through emulation poses many difficulties. Without access to the original hardware and software on which a document was created, it is often impossible to accurately determine how faithful the emulation software is at reproducing an electronic document as it was originally intended.

When tested using emulation software, Robert Creeley's Computer Files had a number of areas or parts of files that were particularly susceptible to errors. The page header or section of a document that precedes the body of a message was frequently garbled and illegible.[15] This could indicate that the emulation program was unable to faithfully interpret the file's software coding or it could be due to its inability to reference missing associated files such as fonts or helper files that would be necessary to generate a true representation. Further study of the files using emulation software highlighted additional areas that were susceptible to errors in emulation such as line breaks, page breaks, and diacritics or special characters.

The problems experienced when using emulation are best illustrated by comparing the rendering of a single electronic file using different software emulators. By comparing emulation results of the same file using Quickview Plus by JASC and Conversions Plus by DataVis, it is immediately apparent that there are significant differences in how a file is visually rendered.[16] The file rendered by Quickview Plus displays garbled or unrecognizable characters or macro encodings in the header. When the same file is emulated using Conversions Plus, the file header is missing, the document's layout or whitespace appears to be misrepresented, and the carriage returns are missing. These differences in emulation are very difficult to quantify as they vary significantly depending upon the type of file being emulated.

This imperfect rendering has a significant impact on the evidentiary value of Robert Creeley's Computer Files to scholars. Does emulation provide a faithful enough representation of the computer files to justify their use in scholarship? Unfortunately there is no easy answer to this question. It is only by making Robert Creeley's Computer Files accessible and

Emulation of schwabsk.y1 file using Quickview Plus by JASC. Reproduced courtesy of Robert Creeley and Stanford University Libraries.

Emulation of schwabsk.y1 file using Conversions Plus by DataVis. Reproduced courtesy of Robert Creeley and Stanford University Libraries.

The Filemaker Pro database used to describe Robert Creeley's Computer Files. Reproduced courtesy of Stanford University Libraries.

assessing the scholarly use of this resource that insights to these questions will be found.

Migration

Migration is the transfer of a computer file from one hardware or software configuration to a newer format.[17] It can be an effective preservation strategy particularly if there is consistency in the file formats being considered for migration. Migration as a preservation strategy for this project was considered but in the end rejected. Robert Creeley's Computer Files contained too many diverse file types (they spanned more than a decade) to make migration a viable approach for this project.

Collection database

Quickview Plus was used to render the files. Descriptive information was then entered into a Filemaker Pro database designed to record file information from the refreshed files. Information such as parent–child relationships between the files, the file names, date information, the number of files, whether or not the files were compressed, and information on file size were all captured. Once this descriptive information was entered, a calculation field was created in the database that concatenated and exported this descriptive information into a well-formed electronic finding aid.[18]

Encoded Archival Description (EAD)

Stanford's long history and experience with EAD made it a natural choice as the preferred method for providing access to Robert Creeley's Computer Files. EAD is a standard developed for electronically representing the relationships between archival materials. It provides a set of guidelines on how to electronically describe archival collections, and a language for standardizing descriptive practices in an open non-proprietary format.[19]

The incorporation of floppy disk images into the finding aid

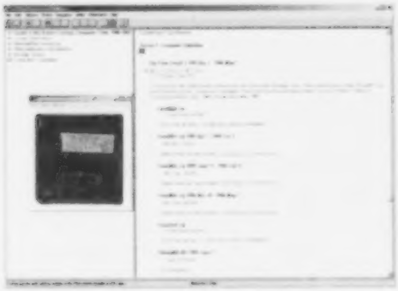
The decision was made to incorporate low-resolution images of each $3^{1}/_{2}$-inch floppy disk into the finding aid on the premise that future researchers and scholars may find the relationship between the storage media and electronic files an interesting topic of inquiry. Providing an image of the storage medium in the finding aid allows researchers insight into how one American poet utilized computer technology over time.

Describing an electronic resource

There are significant differences in descriptive practice between traditional manuscript materials and an equivalent body of born-digital electronic records.

Describing an electronic resource was extremely tedious and far more time-consuming than initially projected. Description was a two-part process as it was necessary to use the Quickview Plus emulation software to render the files before entering the descriptive data into the Filemaker Pro database. Performing both the file rendering and the description on a single computer workstation requires significant technical skill and concentration. And having to use an emulation program and navigate through a structure of hundreds of files also introduces the possibility for greater inconsistencies and errors in the descriptive process.

It also became apparent that the depth of item level description possible with electronic records differs from that for physical manuscripts. To what level or depth of electronic hierarchy does one describe a collection and how can this be balanced with limited staffing resources? Without a reliable mechanism for automating the electronic description process it is impossible to describe every single file, particularly



EAD finding aid of Robert Creeley's Computer Files (collection number M1081). Reproduced courtesy of Robert Creeley and Stanford University Libraries.

when it is not uncommon for a single $3^{1}/_{2}$-inch floppy disk to contain upwards of 500 files. The process of description required a significantly greater investment of time than initially planned. In total, 83 hours were spent describing Robert Creeley's Computer Files using the collection database.

Privacy and intellectual property rights

In accordance with the Stanford's policy on privacy, all of Robert Creeley's Computer Files containing medical records, student recommendations, and certain financial documents were restricted. This meant that it was necessary to view approximately 12 000 files to ensure that the EAD finding aid didn't contain this type of private information. Any files or zipped e-mail archives that contained restricted materials were appropriately flagged in the database and documented as "restricted" in the finding aid. The collection is remarkable for its even distribution of restricted files. This is most likely due to the fact that the floppy disks were created as backups and an effort was made to squeeze as many files as possible onto a single disk. Approximately 15% of the collection contains restricted information that will remain inaccessible to researchers and the public for the foreseeable future.

The intellectual property rights to this collection reside with Robert Creeley. Access to the Robert Creeley Computer Files for research is permitted but publication or commercial exploitation of this collection requires the permission of the rights holder.

Delivery

Phase 1: February – August 2003

Access to Robert Creeley's Computer Files is being provided in two stages. In the first phase of the test bed project, access consists of an EAD finding aid of the collection's contents.[20] The primary goal of this phase of the project is to determine the best strategy for describing born-digital electronic manuscript materials by assessing the level of interest among the patron community to work with this unique resource. Patrons wishing to access the original files during this phase will be provided with hard copy printouts. Direct access to the electronic files will not be provided pending the installation of a dedicated workstation in the Special Collections Reading Room.

During this period, statistical data on the use of the finding aid will be compiled and interested scholars will be asked to provide their insights as to how they envision using the collection. These data will then be incorporated into refining the descriptive practice for phase two of the delivery process and provide valuable insights into how best to preserve and improve access to other born-digital collections.

Phase 2: October 2003 – March 2004

Beginning in the second phase of the project, patrons will be provided with the ability to view electronic files on a dedicated workstation in Stanford's Special Collections Reading Room. The EAD finding aid will be loaded onto a workstation along with a surrogate copy of the collection. This updated version of the finding aid will contain hyperlinks to the files. Clicking on the hyperlink will launch the specified file in the Quickview Plus program allowing the patron to see the rendered electronic record. It is during this delivery phase when intellectual property rights and privacy issues will become a serious concern. It is expected that the development of phase two of the delivery environment will require another significant investment in staff time to adequately ensure that both intellectual property rights and privacy concerns are met.

Conclusion

The preliminary work on this project has emphasized how preserving and developing accessibility to electronic resources requires a significant institutional investment in staffing resources. In total, more than 120 hours have been invested in preserving and describing Robert Creeley's Computer Files. Further analysis shows that 70% of this time was spent creating metadata for the data refreshing process (creating ASCII metadata files) and describing the files and file structure of the collection. It is clear from this project that insufficient metadata remains a principal stumbling block to preserving and providing access to electronic records.

One institution's approach to preserving electronic records should not necessarily be adopted by others. Electronic records are too diverse to mandate that one digital preservation solution fit all circumstances. In the case of Robert Creeley's Computer Files, refreshing the endangered floppy disks onto a more reliable storage medium and emulation were the best fit. This requires that academic and cultural heritage institutions acquire a better understanding of the role of electronic resources in academic scholarship and as artifacts of cultural memory. The application of digital preservation methods such as emulation and migration needs to reflect how collections of electronic records will be used by future generations, and linking access with digital preservation is the best approach to discovering how electronic records will be used.

Stanford's experience with Robert Creeley's Computer Files emphasizes the need for academic and cultural heritage institutions to adopt a proactive approach towards electronic record preservation. These institutions need to invest in sufficiently skilled staff to educate potential record donors and curatorial staff in the benefits of developing recommended protocols for acquiring electronic records.

Endnotes

1. Stanford Special Collections houses two Robert Creeley manuscript collections — Robert Creeley's Computer Files (collection number M1081) and The Robert Creeley Papers (collection number M0662) — which have been maintained as separate collections (electronic finding aids for both collections can be found at www-sul.stanford.edu/depts/spc/spc.html). The files described in this paper are those of collection number M1081.

2. A detailed description of the organizational structure and responsibilities of Stanford University Libraries Preservation Department can be found on their Web site (www-sul.stanford.edu/depts/pres/pres.html).

3. Electronic media in Stanford Special Collections includes large quantities of born-digital media, digitized collections, and audiovisual materials. Robert Creeley's Computer Files is only one of the electronic media collections currently undergoing preservation. For a listing of additional digital collections slated for eventual storage in the Stanford Digital Repository, see their Web site (www-sul.stanford.edu/depts/ts/tsdepts/cat/units/metadata/docs/digi_archive.pdf).

4. More information on the physical properties and preservation challenges associated with computer floppy disks is available in: Ross, S., and A. Gow. *Digital Archaeology: Rescuing Neglected and Damaged Data Sources*. London and Bristol, UK: British Library and Joint Information Systems Committee, 1999, 81 pp. www.ukoln.ac.uk/services/elib/papers/ supporting/pdf/p2.pdf

5. Robert Creeley's Computer Files will eventually be ingested into Stanford's Digital Repository. This OAIS-compliant repository is currently under development.

6. The creation of the Stanford Media Id has been a cooperative effort between Stanford's Principal Manuscripts Processing Librarian, the Media Preservation Librarian, and the Special Collections Assistant Librarian for Electronic Media Materials.

7. In both of these cases, the inability to read the floppy disks could be attributed to an error in the File Allocation Table (which is critically important because it provides an operating system with the file name and cluster information necessary to successfully locate, piece together, and open a file).

8. The Commodore Amiga 2000 computer dates from the mid 1980s. The Amiga file format is unreadable using current Microsoft Operating Systems.

9. Special Collections, Stanford University Libraries. Media Id: M1081_S1_SS0_1_003_OR.

10. Bearman, D., and J. Trant. "Authenticity of Digital Resources." *D-Lib Magazine* (June 1998), p. 1. www.dlib.org/dlib/june98/06bearman.html

11. For an in-depth discussion of the fundamental criteria of authenticity see: Lynch, C. *Authenticity and Integrity in the Digital Environment: An Exploratory Analysis of the Central Role of Trust*. August 2001, pp. 2–3. www.clir.org/pubs/reports/pub92/lynch.html

12. The International Research on Permanent Authentic Records in Electronic Systems (InterPARES) project headed by Luciana Duranti has provided excellent insights into developing a theoretical understanding of issues involved in verifying the authenticity of electronic records. For more information see: Duranti, L. "The InterPARES Project." In *How Do You Know It's the Real Thing? Authentic Documents in the Electronic Age*. Rome, Italy: Ministero per i beni e le attivitáa culturali, Ufficio centrale per i beni archivistici, 2001.

13. Tony Hendley provides an excellent discussion of the application of preservation methodologies to different classes of electronic records. For more information see:
Hendley, T. "Comparison of Methods & Costs of Digital Preservation." In *British Library Research & Innovation Report* (unpublished), pp. 13–14. www.ukoln.ac.uk/services/elib/papers/supporting/pdf/hendley-report.pdf

14. Lazinger, S. *Digital Preservation and Metadata: History, Theory, Practice*. Englewood, CO: Libraries Unlimited, 2001, p. 77.

15. Macros are user-defined programs that can be embedded inside a document. They are often used in word-processing documents to generate automatic date stamps, electronic signatures, and address headers. Macros frequently cause emulation problems because they are often created using vender-specific programming languages.

16. This comparison in emulation is not intended to promote the use of one commercial emulator over another but rather to illustrate the difficulties and differences that occur using emulation. Results in emulation are not scientific and success is highly variable depending upon the format type of the digital object being rendered and the computer hardware running the software. More information on Quickview Plus version 6 can be found on the JASC Web site (www.jasc.com), and more information on Conversions Plus version 6.0, can be found on the DataViz Web site (dataviz.com).

17. Lazinger, S. *Digital Preservation and Metadata: History, Theory, Practice*. Englewood, CO: Libraries Unlimited, 2001, p. 77.

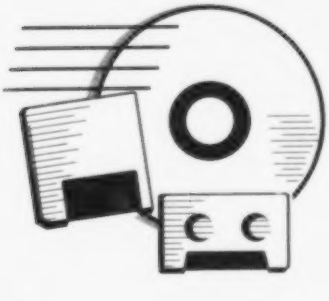
18. Stanford Special Collections makes extensive use of Filemaker Pro databases to record descriptions of all of its collections. Using a database to capture this information allows for easy updating of the collection descriptions and the generation of updated electronic finding aids using a calculation field that concatenates the data into a well-formed EAD.

19. For more information on the EAD standard, see the official Web site (www.loc.gov/ead) hosted by the Library of Congress.

20. The finding aid for Robert Creeley's Computer Files is posted on the Stanford University Libraries Web site (www-sul.stanford.edu/depts/spc/spc.html) where it can be located by the subject of American Literature, the title, or the collection number (M1081).

Résumé

Les dossiers informatiques de Robert Creeley jouent un rôle de premier plan à Stanford : c'est une matrice d'évaluation qui permet d'élaborer de nouvelles stratégies de préservation pour les documents numériques originaux. Dès le départ, il s'agissait d'élaborer une méthodologie intégrant accessibilité et préservation numérique et d'en évaluer l'efficacité. Par accessibilité, il faut entendre le fait de donner la possibilité de découvrir une ressource numérique et de créer les moyens par lesquels les chercheurs et les particuliers intéressés pourront entrer en contact avec cette ressource. Nous soulignerons la façon dont les décisions relatives à la préservation et celles relatives à l'accès sont inséparables et nous verrons qu'elles supposent de mieux comprendre la façon dont les usagers utiliseront les ressources électroniques. Nous aborderons les questions de la dotation, de la migration par opposition à l'émulation, de l'authenticité, des droits de propriété intellectuelle et de protection des renseignements personnels du point de vue de la préservation et de l'accessibilité de cette ressource unique.



OLD WINE IN NEW SKINS: MIGRATING LEGACY DATA IN A MUSEUM CONTEXT

RICHARD GERRARD[1]

Registrar, Museums and Heritage Services
Culture Division
City of Toronto
Toronto ON
Canada
Tel.: (416) 392-6913
Fax: (416) 392-6827
E-mail: rgerrard@toronto.ca

DAVID HARKNESS

Master of Museum Studies
Candidate
Museum Studies Program
University of Toronto
Toronto ON
Canada

Abstract

This paper presents an extended case study describing the evolution of the digital collections management systems at Museums and Heritage Services, City of Toronto, over the last 15 years. This experience likely echoes the information technology adoption experience of many smaller museums. The trail begins with stand-alone personal computers, moves to workstations in a networked environment, and progresses finally to Web-based Inter/Intranet applications.

Introduction

The City of Toronto began keeping paper-based records in 1881 when it acquired its first museum, Colborne Lodge. More than a century later the City of Toronto museums started to replace their paper systems with various collection management systems on stand-alone personal computers. Since that time the collection data have been migrated through at least nine different custom and commercial software packages.[2] The legacy is multiple individual datasets and file formats stored on a wide range of physical media types.[3]

This paper presents an extended case study describing the evolution of the digital collection management systems from 1987 to 2003 at the museums that comprise the Museums and Heritage Services Unit, Culture Division, City of Toronto (MHS). [This Unit was assigned the responsibility for managing the museum collections of the former municipalities of York, East York, North York, Etobicoke, Scarborough, Metropolitan Toronto, and the Toronto Historical Board in 2000.] This experience likely echoes the information technology adoption experience of many smaller museums. The trail begins with stand-alone personal computers, moves to workstations in a networked environment, and progresses finally to Web-based Inter/Intranet applications.

After 15 years of migrating databases, two distinct scenarios driving data migration projects have been observed: the straight swap and the re-engineering migration. The straight swap is usually initiated by the need to "upgrade" and results in few changes other than the underlying software. In the case of a re-engineering migration, the need to add new functionality to the software is the catalyst to change the system. The common threads in these two scenarios are the need to retain the existing digital data in order to preserve the significant investment that has already been made in compiling, editing, and managing this resource, and the desire to make the transition as transparent as possible in order to minimize the impact on the ongoing business of the museum operations dependent on the data.

Museum registration: The power of paper

Paper-based systems have not disappeared. And to better understand the digital environment, it is necessary to first have some idea of the paper-based one that preceded it.

The earliest systematic record-keeping in museums began with acquisition ledgers.[4] These documents were basically "laundry lists" of all objects entering and leaving the possession of the museum, as well as data describing who owned or donated them, their insurance value, where they were physically located, and details about the object itself. [One of the City of Toronto museums, Historic Fort York, still possesses an acquisition ledger dating to its opening in the mid 1930s.] Acquisition ledgers were later augmented with additional documents (e.g. inventories and catalogues, books and other scholarly publications) and object-level descriptions on the catalogue cards. In fact, collection data have grown significantly in complexity — the result of creating additional specialized document types as new professional specialities and technologies were brought to bear on the collection and its objects.[5]

Very personal computers: Stand-alones with homemade applications

It was in 1987 that the Toronto Historical Board (THB) received its first microcomputer, an IBM-PS2 model 50 equipped with dBaseIII+ that landed in the Fort York Archaeology Programme's material culture lab. One of this paper's authors (Gerrard) was instructed to use this new tool to create a database to catalogue the 70 000+ specimens that had been recovered during the Programme's first-season excavation. The result was the *Fort York Archaeology Programme Material Culture Data Management System*.[6] The next task was to migrate the existing record sets from the old data storage systems (i.e. paper catalogue worksheets) into the new data storage system (i.e. digital catalogue worksheets), thus (re)creating the same paper-based system on the computer.

The first application of this experience to the THB's museum collection was in 1989. The development approach employed was simple and direct: use the existing functional model (i.e. the paper-based museum registration data and procedures) and replicate it in a digital format. Like nearly all Canadian museums at the time, the THB employed the Humanities Data Dictionary developed by the Canadian Heritage Information Network (CHIN) to guide its cataloguing work.[7]

Two limitations of this system quickly became apparent. First, because the application was situated on a stand-alone personal computer (PC) sitting on the registrar's desk, external users, curators, and other staff were unable to access the collection directly — they still had to telephone the registrar's office, request information about a certain object, and wait until it was located online or in the physical records. Second, the technology was unable to permit complete transfer of the richness of the information located in the physical records into the computer database. Due to these limitations, it was still necessary to maintain the extensive (and expensive) paper-based system in parallel with the new digital database. In practice, the database was (and is) relegated to the function of a simple finding aid for the paper records; it provides cross-indexing functionality hitherto limited to multiple card indexes or staff memory.

A new network of relationships: Networks and relational data models

In the early 1990s the THB migrated its flat-file databases from stand-alone PCs to a fully relational database management system (DBMS) in a networked environment. The goal here was two-fold: first, to improve data management and query efficiency; and second, to develop the ability to share the database among a wider user base. Like most information technology (IT) projects the results were mixed.

The THB did improve its efficiency in both data management and query capability when the data were transferred from dBase to Oracle in 1990.[8] In addition, as part of the migration project, the THB decided to continue development of the database's capabilities. For example, several object description tables were added to the limited accession ledger data. This permitted specialized information related to specific portions of the collection (e.g. furniture, books, fine art) or new operational requirements (i.e. conservation) to be stored, manipulated, and retrieved.[9] The THB was less successful in creating a wider user base. The ability to use the Structured Query Language (SQL) greatly improved the efficiency of searching the database, but few users outside the registrar's office wanted to learn SQL programming to do their own searches. Oracle's query-by-form (QBF) capabilities were also crude and unreliable at that time, so collection management staff remained the primary users of the system.

In 1995, the THB again tried to improve user access to the database. First, MS Windows was adopted as the corporate standard operating system. Because Oracle was not compatible with MS Windows at the time, the database had to be migrated to another database application. After testing several different database packages, MS Access was selected. However, it should be noted that the decision was not based solely on an analysis of the product. It was heavily influenced by the THB's decision to adopt MS Office as an office automation suite. With the database on the corporate network and using more "user friendly" software, the THB hoped to overcome some of the limitations experienced by users.

The immediate technical challenge was to migrate and merge nearly 20 individual data tables into a re-engineered data structure. This new structure was designed to place all the collection records into a single searchable database instead of the multiplicity of specialized databases.[10] As might be expected, several problems were encountered when trying to move the data contained in one proprietary database file type (Oracle) to another (MS Access).[11] These problems were due, at least in part, to a lack of experience and understanding about the number of ways digital data can get mangled or lost during the simple act of moving to another software. Fortunately, the post-migration quality control was assisted by two important facts: first, the staff involved with the data migration project were the same individuals who had undertaken the physical inventories that created the digital records in the first instance; second, the record sets were relatively small (usually less than 1000 records per set). This detailed knowledge of the collection plus the ability to deal with the data in small, coherent pieces allowed staff to scan for, identify, and correct errors in the data almost immediately. Unfortunately the outcome was similar to the THB's previous migration: the system got more sophisticated and the user base did not get any broader.

Organizing for the Web: Virtual Museum of Toronto

In 2001, following the amalgamation of the City of Toronto and the creation of the Culture Division's Museums and Heritage Unit (MHS), staff and management began a series of discussions about the creation of a Virtual Museum of Toronto (VMT). The initial concept was to place the collection database online as a first step towards the VMT. However, after examining how the information was organized in the database and various other external factors, the MHS team concluded that simply mounting the collection database was not going to meet public access requirements. What was required was re-engineering of the operating model that assumed digital data were being accessed by staff with expert knowledge, and re-examination of how the data themselves were contextualized for public use. In short, the paradigm that had been used for the collection database during the past 15 years needed to be fundamentally altered.

We looked to the *Rules for Archival Description* for inspiration and began by developing a top-down descriptive model of the entire collection.[12] By adapting archival concepts like fond and series, rather than material or function, for the higher-level descriptive units,[13] we hoped to rapidly deploy a significant portion of the collection onto the Web site.

This model worked well until it was applied to the Larry Becker collection, a large collection of numismatics and memorabilia. This collector had divided the 4500+ objects in his collection into about 20 sub-collections, and then physically organized the objects into 140 specimen trays that cross cut these sub-collections. The MHS wished to preserve this original order to demonstrate the collector's curatorial vision of his collection, and to make the collection publicly available through the City of Toronto Web site. Discussions about how to achieve these goals raised several issues about the newly developed Web data model that required resolution.

First, how would it be possible to maintain the original physical order in the database? The solution was simple: treat the objects in each specimen tray as if they were analogous to an archaeological assemblage from a specific excavation unit. Implementing this required the addition of a new field (called TRAY) to the database at the object level, and to record the tray number there.

Next, what was the best way to deal with the collector's subject-level sub-collections? This proved to be more difficult. As with most subject-level cataloguing, the classes were not mutually exclusive. One object might easily fit into several sub-collections.[14] The solution was to fall back on the donor's original order. Again the technical solution was to add a new field (called SUBCOLLECTION) to the database to contain the description of the sub-collection.

The most difficult decision became trying to place TRAY and SUBCOLLECTION into the hierarchical model. Which should be made equivalent to a series in the collection-fond-series-object model? Both are composed of individual objects, so which one to choose? The MHS team eventually chose SUBCOLLECTION as the series, and were then left with the inelegant work of defining each specimen tray as a composite object consisting of a number of individual objects.[15] But the original simple hierarchical model was not designed to accommodate these alternative methods of grouping objects; objects could not be placed into more than one series, and creating groups of objects from several series was, by definition, impossible.

A new beginning

In 2001–2002, a team of MHS and City of Toronto IT staff[16] worked together to develop GORMS (Generic Object Repository Management System), a database solution to allow the storage of museum multimedia object files in a structured format to permit public access via the City of Toronto Web site. GORMS is a digital assets management system and departs from traditional museum collection management models in many respects. However, it begins to resolve some of the issues raised above about creating digital collections. The GORMS database is about managing the digital surrogates of physical objects — jpg and bmp files — rather than the objects themselves,[17] but by including artifact descriptions as part of the metadata it can be used to provide online access to a limited set of descriptive collection data.[18]

More importantly, GORMS deals with the need to create meta-collections that exist outside the formal classification systems used to describe the collection. It accomplishes this in the following way. The "Container" entity represents a recursive, hierarchical structure in which the objects are organized. The IT staff viewed this as being similar to the "folder" concept from the MS Windows operating system. However, the structure is much more flexible — allowing the user to develop multiple container hierarchies and to assign objects to several separate containers simultaneously.

Linking the "Object" and "Container" concepts is a "Collection Object" entity. This logical representation acts as the digital analog to the specimen tray or composite object. "Collection Objects" can contain multiple digital objects and in specific user-selected sequences. This is the solution to the Becker Collection trays problem discussed above. The GORMS structure allows multiple nested sets to be created from individual digital objects, and for these to be annotated and stored for future reference in the database. By using the same structure, it is also possible to implement multiple formal classification systems to organize the digital objects in the collection.

Data migration: Process and pitfalls

The Todmorden Mills Museum was amalgamated into MHS from the former municipality of East York in 2000. In 2002, after nearly 15 years of service, their Heritage Sentinel collection database was migrated to the MHS in-house collection database. This required translating the data from the now obsolete proprietary software format to MS Access.[19] The process proved to be complicated for many reasons[20] but, in spite of the technical challenges, it was generally successful.

During this project, and from our previous experience, four distinct classes of data errors (translation, transposition, truncation, and wild card) arising from two different sources (machine error and human error) were encountered. However, it must be emphasized that the data migration projects described in this paper include primarily text, date, or numeric data types residing in table-based databases (flat-file or relational). Additional problems may have arisen had there been image, video, and sound files, but these data types would still be subject to the four error classes discussed below.

Translation errors result in data getting lost or garbled as they go from point A to point B. Changing an attribute's data type (e.g. moving from a text field in the old system to a numeric one in the new) may cause problems in translation for any non-numeric characters.[21] Migration of a date field from one software to another might fail depending on how the date is calculated internally by the respective software or operating systems. The interpretation of character sets may generate incompatibilities ranging from fairly benign (non-supported characters translated into garbage) to devastating (control characters being rejected or misinterpreted). Some of these software-based errors can be controlled if they can be identified during the planning stage, but more often they appear only during the first attempt to read the data following migration. This turns the migration into an iterative process, with each successive generation resulting in a cleaner dataset as errors are discovered and corrected.

Transposition errors are largely a human creation. In these cases, data are not necessarily lost but instead become "misplaced" within the new database. These simple errors most often occur while mapping the old attributes onto the new. Other transpositions can occur when concatenating several attributes into a single attribute (in which case the data might all be there but could be in the "wrong" order), or inversely when parsing the data in a single attribute into several new attributes, e.g. if original data have multiple attributes in a variable order (e.g. length-width-height in one record and height-width in the next) it is easy to place data into the incorrect field at the time of parsing.

Truncation errors can occur when too many data are squeezed into too small an attribute size.[22] For example, when exporting data from an MS Access Memo field to an MS Excel cell, data can be truncated at 255 characters. A variation on this is truncating the number of fields to the maximum limit of the new database software (e.g. MS Access has a 255 field per table limit). This is the result of system limitations combined with human error. Many of these problems can be anticipated, but it often requires a very detailed knowledge of the software being used and the dataset involved. And unfortunately solutions are not always simple or obvious.

The final type of errors have been categorized as "wild cards." These have the annoying habit of being both random and expected with digital data. They are both human and technological in origin, and often strike without warning. They can manifest themselves as a catastrophic hardware failure such as a shredded magnetic tape or hard disk drive crash, or they can be more subtle cases of "bit-rot"[23] (the random bit losses that accumulate on magnetic media until a file becomes unreadable[24]). Ultimately the data may be recoverable, but the Herculean efforts required are often beyond most museums to undertake with in-house resources.

Final thoughts

After a decade-and-a-half it is possible to observe two distinct scenarios for data migration projects. The first is the straight swap in which there is little or no change to the data or business model, although the underlying software supporting them gets updated. Our first move from paper to bits is an example of this type of migration. Similarly the movement of the Todmorden Mills Museum database from Heritage Sentinel to the MHS main collection database changed little about the data while completely updating the underlying technology. This is possible because they are based in the same underlying data dictionary and procedures. The second scenario is a re-engineering migration. In these cases the need to add new functionality to the software is the catalyst to change the system. Moving from stand-alone workstations to a network environment, or from a network environment onto the Web, are good examples of this in action.

It is only through careful planning and execution of the migration process that the majority of errors can be avoided. However, post-migration quality control procedures are also required to identify and correct errors that are introduced during the process. Additional data management strategies (e.g. backup and recovery plans, multiple copies of mission critical data, and migrating data from old to new storage media) are required during the migration process to ensure the long-term preservation of digital assets.

A cost–benefit analysis should be conducted before undertaking migration of a dataset from an old(er) format to a new(er) one. Some of the questions that should be considered include: Is the cost of simply re-entering the data into the new database more or less than the cost of migrating the data directly from the old database? Do you have still have software, hardware, and human resources that understand the legacy system? How many records are there? How clean are the data? How complex are the data? What is the current error rate in the data? How many and what type of new errors might be introduced during the transfer?

In spite of everything that has been discussed in this paper, in general we feel that most digital collection data are worth migrating (or at the very least archiving). This is because, as with old paper collection records, one can never be sure that buried somewhere in the mass of records there is not one irreplaceable scrap of evidence vital to resolving a future collection management problem.

Endnotes

1. Author to whom correspondence should be addressed.

2. These include dBase, Heritage Sentinel [Advanced Revelation], ORACLE, SuperBase, WordPerfect, Lotus 1-2-3, MS Excel, MS Access, and SQL-Server.

3. This includes 3.5-inch floppy disks, various Winchester-type hard drives, Iomega Ditto drive, four different magnetic tape formats, videodisk, and several CD formats.

4. There are "how-to" collections management guides published as early as the mid 16th century. A format nearly identical to modern acquisition ledgers has been used at The British Museum since the mid 18th century. See Gerrard (1994, 1996) for a discussion on these early collection management systems.

5. The documents may be derivative of or summarize other documents with specific expert interpretation / commentary as added value text. In addition, many of these documents are analogous to earlier document types, illustrating a succession as one technology supplants another (e.g. photographs being the visual equivalent to the earlier artists' renderings).

6. See Gerrard (1988) for a description of the application.

7. See Delroy et al. (1988).

8. The THB ran Oracle on stand-alone workstations running under the MS DOS operating system. See Gerrard (1989) for a description.

9. For a description of the system see Gerrard and Knowles (1990).

10. These separate datasets were created as part of individual physical inventory projects. Each dataset consisted of a core series of attributes augmented with descriptive attributes specific to the object class, therefore the furniture, fine art, book, marine history, weapon, military uniform, tools, housewares, and archival collections all possessed slightly different attribute sets.

11. In addition, we had several inventories created as object lists using various word processing formats (e.g. WordPerfect) and a few in spreadsheet applications (e.g. Lotus 1-2-3).

12. See www.cdncouncilarchives.ca/rad_ch1.pdf for a description of RAD.

13. Our initial hierarchy was collection-fond-series-object. See w3.toronto.ca/culture/virtualmuseum.htm for the implementation of the structure.

14. For example, where do you place a Canadian National Exhibition award medal presented to Sir Henry Pellatt (builder of Casa Loma in Toronto, Canada) and placed by Mr. Becker into a tray of medals relating to the history of the Queen's Own Rifles of Canada (Pellatt was their Colonel-in-Chief) when two of the sub-collections are Canadian National Exhibition memorabilia and Militaria?

15. The concept of a composite object was developed to describe a set of individual objects that together form a culturally significant whole. For example, a tea service is a composite object consisting of a teapot and lid, sugar and lid, creamer, and slop bowl as individual objects.

16. This team included Tammy Milekovic, William Hong, Ryan Manuel, and John Liu of the City of Web Management Services, Information and Technologies Division; and the authors, Elisabeth Joy, Lynn Kurylo, and Carl Benn of Museums and Heritage Services, Culture Division.

17. This cultural differences around the definition of the word "object" led to some initial confusion between the IT and museum portions of the development team: the museum staff used "object" to mean the large vase in the conservation lab whereas the IT staff used "object" to refer to 1996.21.234_vase_fullview.jpg, an image file of the large vase.

18. For a technical overview of the data model, see Hong (2002).

19. A full discussion of the technical process is beyond the scope of this paper. For a description of the process, see Harkness (2002).

20. First, there was no documentation of any sort describing Heritage Sentinel or Advanced Revelation, nor did we have the original installation disks. Second, Heritage Sentinel was running on an 8-year-old 80486-based PC; the backup mass storage device was an obsolete internal Iomega Ditto drive. Third, it was running in the MS DOS emulation under MS Windows 95. Fourth, the design of Heritage Sentinel was highly relational with dozens of data and join tables. Each needed to be identified and documented before we could start. Fifth, the only available export formats were ASCII, dBase, and Lotus 1-2-3. Sixth, there was no one on staff trained to administer the system. Finally, after 2 decades, the data contained in this old, obsolete, and cranky database were probably richer than the existing paper record, and for some of the information (e.g. object location) probably not recorded elsewhere. After the equally ancient laser printer ceased to function, even printing a hard copy became a struggle. Even if we could get a new printer to work on the PC, learned the bundled report writer program, and printed off the more than 6000 records in the database, we would be left with a significant data entry and editing task.

21. Microsoft's TechNet Web pages are an excellent resource to begin to identify and resolve these issues. See *ACC2000: Memo Field Truncated When Report Is Output to Excel* (support.microsoft.com/default.aspx?scid=kb;en-us;208801) and *XL: Unexpected Behavior with Numbers Formatted as Text* (support.microsoft.com/default.aspx?scid=kb;en-us;141765) for just two examples.

22. See *XL: Passed Strings Longer Than 255 Characters Are Truncated* (support.microsoft.com/default.aspx?scid=kb;en-us;105416). MS Access has specific limits about the length of a text field (255 characters) and the number of fields in a database (255) among others. See *ACC2002: Access Database and Project Specifications and Limits* (support.microsoft.com/default.aspx?scid=kb;en-us;302524) for details.

23. "Bit-rot" is a term coined by Bruce Sterling in his 1997 novel *Holy Fire* (p. 31).

24. Bit-rot is slowly rendering digital files ranging from Brian Eno's earliest digital recording to NASA and the Pentagon's datatapes unreadable. See MacLean and Davis (1998) for a discussion of the issues.

Bibliography

Delroy, S.H., D. Jewett, and R. Bellamy. *Humanities Data Dictionary of the Canadian Heritage Information Network, Documentation Research Publication Number 1, Revision 2.* Ottawa, ON: Communications Canada, 1988.

Gerrard, R. *Fort York Archaeology Programme Material Culture Data Management System (Version 1.3).* Unpublished report on file with the Fort York Archaeology Programme, Culture Division, City of Toronto, ON, September 1988.

Gerrard, R. *Collections Management Information Systems Framework: A Summary Report.* Unpublished report on file with Collections and Conservation, Culture Division, City of Toronto, ON, 1989, 13 pp.

Gerrard, R. "Evolution, Revolution, and the Laundry List?: A Personal Perspective on Change in Collections Management." *Ontario Museum Annual* 3 (1994), pp. 40–45.

Gerrard, R. "Stewardship, Collections Management and the New Technology: Old Problems, New Challenges." *Heritage Forum*, 1996. www.chin.gc.ca/

Gerrard, R., and M.A. Knowles. "Lessons Learned from ORACLE; The Development of a Collections Management System at the Toronto Historical Board." *Spectra* 17, 3 (1990), pp. 1–8.

Harkness, D. *A Case Study in Heritage Sentinel Data Migration.* Unpublished report on file with Collections and Conservation, Culture Division, City of Toronto, ON, 2002.

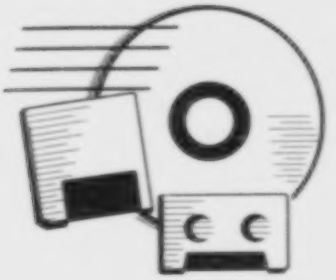
Hong, W. *Generic Object Repository Management System: Database Design Overview, Version 1.0.* Unpublished report on file with Collections and Conservation, Culture Division, City of Toronto, ON, 2002, 33 pp.

Sterling, B. *Holy Fire.* New York, NY: Bantam Books, 1997.

MacLean, M., and B.H. Davis (editors). *Time and Bits: Managing Digital Continuity.* Santa Monica, CA: Getty Publications-Long Now Foundation, 1998.

Résumé

Nous présenterons une étude de cas approfondie qui décrit l'évolution des systèmes numériques de gestion des collections aux Services des musées et du patrimoine (Museums and Heritage Services) de la Ville de Toronto au cours des 15 dernières années. Nous croyons que l'expérience de la Ville dans ce domaine est à l'image de celle de nombreux petits musées. Des ordinateurs personnels autonomes, les données sont passées sur des postes de travail en réseau, puis finalement sur des applications Web Internet et intranet.



A STRATEGY FOR ARCHIVING WEB SITES AT LIBRARY AND ARCHIVES CANADA

SUSAN HAIGH

Senior Policy Officer
Library and Archives Canada
395 Wellington Street
Ottawa ON K1A 0N4
Canada
Tel.: (613) 992-2464
Fax: (613) 943-1939
E-mail: susan.haigh@nlc-bnc.ca

ROSELYN LILLENIIT

Senior Project Officer
Government On-Line Task Force
Library and Archives Canada
395 Wellington Street
Ottawa ON K1A 0N4
Tel.: (613) 996-7493
Fax: (613) 991-9871
E-mail: roselyn.lilleniit@nlc-bnc.ca

Abstract

In 2001–2002, the National Library of Canada (which has since joined with the National Archives of Canada to form Library and Archives Canada) acknowledged that it must play a leading role in ensuring that significant Canadian Web sites are archived and accessible to Canadians over time. It formulated a two-pronged approach to increase its understanding of the requirements, mechanisms, and challenges involved in archiving Canadian Web sites. The approach called for a small pilot project for the negotiated acquisition of a modest number of selected Canadian sites, coupled with joining an international consortium of several national libraries and the Internet Archive to collaborate on developing tools and expertise to build large-scale national heritage collections of Internet content on an automated basis. Progress on both of these initiatives is outlined, as well as likely next steps.

Introduction

The National Library of Canada (NLC) has been building its online electronic collection for almost 10 years. Following an internal study carried out in 1993, NLC conducted a pilot project from 1994 to 1995 to acquire, preserve, and catalogue a small number of Canadian electronic journals. The Final Report[1] of the pilot project and its summary[2] were published in 1996. The catalogued titles began to be listed in *Canadiana, the National Bibliography*[3] in January 1997, and later that year the Electronic Publications Acquisitions Section was created. In 1998, NLC developed the *Networked Electronic Publications Policy and Guidelines*.[4]

The Electronic Collection consists of Canadian books and periodicals published online. As of December 2003, it includes more than 11 451 titles and more than 46 000 serial issues published by the commercial publishing sector and the government publishing sector (with considerable collecting emphasis placed on the government sector). The archived publications exist in the format in which they were produced, including html, pdf, ASCII, and others. Archived online publications are catalogued in the same manner as printed publications, and bibliographic records are available via AMICUS, the national union catalogue. Some titles have restricted access, and are accessible through NLC staff or by using the public Internet workstations in the NLC Reference Room.

While these developments were taking place, another project involving Statistics Canada publications came to fruition in 1996. This project was headed by Statistics Canada and the Depository Services Program, and NLC took part as a depository library. The results of this experience are presented in *Electronic Publications Pilot 1996/97: Final Report*.[5]

Historically, responsibility for collecting Canadian material has generally been divided between NLC and the National Archives of Canada (NAC) based on whether or not the material was formally published: NLC collected published Canadiana, and NAC collected unpublished Canadian material of national significance as well as the records of the federal government. Thus, in the late 1990s, these two entities discussed informally that NLC might bear responsibility for Internet sites, while NAC would be more appropriately concerned with Intranet sites, e-mail, and electronic records. Neither, however, was ready to routinely collect such material.

In 2001, the National Archivist was asked to archive the Government of Canada's Millennium Bureau Web site, and also agreed to acquire and preserve the site of the private Our Millennium Foundation.

In 2002, NAC and NLC were both approached about archiving the Web site of the Commission on the Future of Health Care in Canada. This latter site was scheduled to be removed from the Internet at the end of 2002, and all parties were concerned that it would become inaccessible to Canadians at the very time that interest in it would be at its highest.

These sites are representative of the many valuable Canadian Web sites that have been moved or discontinued and to which, without intervention to archive, access will be lost.

In 2001–2002, NLC [which has since joined with NAC to form Library and Archives Canada (LAC)] acknowledged that it must play a leading role in ensuring that significant Canadian Web sites are archived and accessible to Canadians over time. It formulated a two-pronged approach to increase its understanding of the requirements, mechanisms, and challenges involved in archiving Canadian Web sites:

- a small pilot project for the negotiated acquisition of a modest number of selected Canadian sites
- participation in an international consortium of several national libraries and the Internet Archive to collaborate on developing tools and expertise to build large-scale national heritage collections of Internet content on an automated basis

A third very important enabler, a clause that will give LAC the right to harvest Internet content, has also been included in the legislation now before Parliament to create Library and Archives Canada. These initiatives are detailed below.

Web site archiving pilot project

In December 2002, the Web Archiving and Preservation Committee at NLC approved a proof-of-concept project to negotiate with specific Canadian Web site owners to acquire a modest number of selected Canadian Web sites for inclusion in the Library's Electronic Collection. The project supported two strategic priorities: preserving and protecting the national collection, and providing free access for all Canadians to knowledge on Canada. A small-scale project of this nature would inform us about Web site owners' willingness and concerns regarding archiving and access, and would allow us to explore legal issues, metadata, copyright requirements, information retrieval issues, and technical and preservation challenges. This project would also provide an opportunity for staff to become more informed and ready for productive participation in the Internet archiving consortium of national libraries beginning in 2003.

Proposed project approach

Planning for this brief project began before Christmas 2002, and the project ended April 30, 2003. The preparation of an explanatory letter/e-mail to send to the Web site owners plus a form for copyright clearance for materials on each Web site was identified as a major requirement.

The Electronic Publications Pilot System (EPPS), which is currently employed by the Electronic Publications Acquisitions Section to acquire electronic publications, was chosen as the management tool for archiving the selected Web sites during the pilot project. Details about the number of snapshots to be captured over a specific period of time would be determined during the negotiations with the various Web site owners.

The team decided to apply procedures and workflow analogous to those currently followed to acquire electronic serials for the NLC Electronic Collection, e.g. appropriate metadata would be created in the EPPS and acquired Web sites would be catalogued in AMICUS, the national catalogue. The goal was to ensure that Canadians would be able to access the archived Web sites via the Electronic Collection on the NLC Web site and also by searching in the AMICUS database. Any restrictions regarding access imposed by Web site owners would be dealt with on a case-by-case basis. All unique needs, differences, problems, and concerns would be documented as they were encountered.

The pilot project aimed to help determine a workable "blueprint" that LAC could use to expand significantly the acquisition and archiving of Canadian Web sites starting in 2003–2004.

Selection of Web sites

The number of Web sites to be archived was forecast to be about 20 or less. The Web site list used in the "domain.uk" project as well as the National Library of Australia's selection of Australian Web sites for inclusion in PANDORA[6] served as models for the selection of the Canadian Web sites. Consultation with staff in the NLC Reference Division helped to identify appropriate Web sites. A popular NLC Web service, Canadian Information by Subject (CIBS), was also used to identify possible candidates.

Types of sites considered for the pilot project included:

- government Web sites (federal, provincial, territorial, municipal)
- commercial Web sites, especially those related to publishing or book trade

- defunct, or soon to be defunct, Web sites, especially government ones
- Web sites for authors, musicians, literary figures, etc.

Types of Web sites excluded from the pilot project included:
- databases that required the input of a search term from a user
- very large Web sites
- sites where the copyright status of music, lyrics, or images was uncertain
- sites for which considerable work would be required to obtain copyright clearances
- sites with password or registration requirements
- personal sites of individuals (as noted above, this exclusion did not pertain to cultural figures such as authors, musicians, etc.)

Based on the above criteria, the project team researched and compiled a list of approximately 50 suitable Web sites.

Participation of Canadian Web sites

Early in January 2003 an explanatory letter/e-mail and a form for the copyright permission required for participating in the archiving Canadian Web sites pilot project were drafted. Over the course of the next few months, these were sent out to the owners of the following eight Web sites (a brief indication of the owners' responses and the issues that arose are included in parentheses):
- Media Awareness (positive response; third-party copyright problems)
- GovNet Conference (positive response; third-party copyright problems)
- Prime Minister's site at Privy Council Office (positive response; work is still in progress)
- Polaris Institute (positive response; SuperBot software changed absolute links to relative ones; arrangements were made to receive the files via FTP, however, one sub-directory was missing; follow-up continues)
- Canadian Sports Hall of Fame (positive response; however, owner subsequently lost his files due to an error on the part of his Internet Provider; work continues)
- NatureWatch [composed of PlantWatch, FrogWatch, WormWatch and IceWatch] (positive response; problems occurred with proprietary file formats, e.g. asp files used by one of the sub-sites; this was a common problem with many Web sites considered for inclusion in the pilot project)
- Canadian Nature Federation (e-mail request bounced back, i.e. was not deliverable)
- WorkRights (no response to e-mail request)

Issues revealed

The small team that worked on the project concluded that there were three main issues:
- Copyright clearance — Some Web sites mount material owned by third parties and obtain clearance for this material to be used only on their specific Web site. In such cases, LAC felt it was unable to archive the site because permission had not been explicitly granted by rights owners for storage and public access on LAC servers.
- Software tools — Project personnel spent a considerable amount of time searching for and testing software tools that would download Web sites consistently and accurately and which could convert proprietary format files such as asp to html. A tool that would work well within the EPPS environment was not found within the project time frame, although several (SuperBot, Web Whacker, Grab-a-Site) were tested.
- Technical issues — Several of the sites selected could not be downloaded or made accessible with the tools used due to the use of certain file formats within the sites. Some of these formats pose security concerns or did not operate on the existing NLC Web server, which uses a Unix operating system and Apache Web server software. In our present environment, the following file formats could not be acquired due to security concerns, server restrictions, and search requirements: asp, shtml, css, hhtml, xml, ilml, ht access files, ssi, php. Loading the Web sites on a separate server (asp requires the Microsoft IIS Web server software) was not desirable because this approach would require that the NLC Web server keyword searching set-up be changed to permit searching across more than one server. Technical resources to make these changes to the keyword searching set-up were not available during the time frame of the pilot project.

The use of proprietary file formats on Web sites severely limited the number of Web sites that could be considered for inclusion in the pilot project. For example, the Web sites of the major Canadian political parties (e.g. the Liberal Party, Canadian Alliance, and the Conservative Party) use these types of file formats and therefore they could not be contacted and invited to participate in the project. In contrast, the Web sites of many smaller political parties (e.g. the Marihuana Party, the Green Party, and the Communist Party) use simpler more standard file formats and were therefore possible candidates for inclusion. However, to avoid an imbalance in coverage, no political party Web sites were included in the pilot project.

Project evaluation and required follow-up

Due to the issues described above, none of the Web sites had been successfully archived and made accessible to our clients by the end of March 2003. Given these technical constraints, the review committee decided that emphasis should first be placed on capturing certain Web sites on our servers. After capture, further testing could be undertaken to determine how best to provide access to the Web sites for our clients. The two sites that have been transferred to our custody but are not yet available through LAC are both from the Privy Council Office: the Health Care Commission's site and the Prime Minister's site.

This small pilot project was useful in pinpointing the areas on which LAC could focus to be better equipped to acquire and process Canadian electronic resources. We learned, for example, that many site owners support the idea of their sites entering a national archive but they would like the captured site to be clearly identified as the archived version. It was clear that LAC needs to analyse its technical environment and make the changes necessary for handling Web sites as part of our collections, and it must designate staff to both investigate software tools and to undertake the work of selective acquisition more effectively. The project made us aware of the constraints we are facing with this challenge: we don't have the infrastructure or human resources necessary, and there are barriers as far as securing permission from Web site owners that also impede our progress.

The project underscored that negotiated acquisition of selected Web sites through file transfer is time-consuming and labour-intensive, and not scalable to the capture of a large representative sample of Canadian Web sites. This leads us to the second prong of LAC's strategy, collaborating with other national libraries to explore how to build large collections by more automated mechanisms.

Internet preservation consortium

Background

Initially proposed in 2002 by the Internet Archive at IFLA[7] and at the Sixth European Conference on Digital Libraries' Web Archiving Workshop,[8] the consortium was conceived as a way to foster international collaboration among major national libraries to collectively define collection policy and common practice for Web archiving. The consortium would specify and develop a new Web crawler targeted towards the archival needs of national libraries. The Internet Archive would provide a national collection for real test and evaluation.

This notion was well received by those national libraries that have been most active in Web archiving to date. Both the United Kingdom and France[9] had undertaken projects targeting capture of Web sites of various types to explore feasibility and challenges. Denmark had captured sites related to their latest municipal elections, and were planning a three-part program of one event-based crawl per year, four crawls of the .dk domain per year, and continued annual growth in their selective acquisition of online publications and new media. Sweden, the most advanced in this area to date, is currently capturing one or two snapshots per year of the .se domain plus Swedish .com sites and other sites of Swedish significance.[10] Australia, like Canada, had a selective, negotiated approach to collecting mainly online publications as well as some Web sites.[11]

Many of these projects have identified as foremost among their challenges:

- a robust and flexible crawler[12]
- a supportive legislative framework
- an uneven ability to provide access to captured content (some provide no access yet; others allow research access only from national libraries' reading rooms)
- an inability to capture the "deep Web" (databases, passworded sites)
- a difficulty with some file formats (e.g. Flash)
- an uncertainty as to scalable approaches to metadata

Consortium membership and structure

Since that time, several meetings among initial consortium participants have helped solidify an approach and a governance structure for the consortium. The consortium principles and objectives are as follows:

- to work collaboratively, within each country's legislative framework, to identify, develop, and facilitate implementation of solutions for selecting, harvesting, collecting, preserving, and providing access to Internet content
- to facilitate international coverage of Internet content archive collections within national legal frameworks and in accordance with individual national collection development policies
- to be a strong international advocate for initiatives and legislation that encourage the collection, preservation, and access to Internet content

To achieve this aim, the consortium will:

- provide a forum for the sharing of knowledge about Internet content archiving

- develop and recommend standards for collecting, preserving, and providing long-term access to Internet content
- facilitate the development of appropriate and interoperable tools
- raise awareness of Internet preservation issues and initiatives through activities such as conferences, workshops, training events, and publications

Consortium membership is open to any national library or heritage institution that is prepared to actively contribute expertise and/or funding toward either development projects or working groups that are approved by the steering committee. The following libraries have joined:

- Bibliothèque nationale de France (lead)
- British Library
- Library of Congress
- national libraries of Norway, Sweden, Denmark, and Iceland
- Library and Archives Canada
- National Library of Australia
- National Library of Italy
- Internet Archive

Proposed initial tool development project — "Smart" crawler

Work has also begun to define requirements for a sophisticated open source crawler that will be able to capture millions of pages on a timely basis. It must be modular and flexible so that it can be modified and extended over time. It will be customizable so that countries can capture whatever portion of their Web space they specify (e.g. it will support broad or selective crawls, continuous or periodic crawls, thematic crawls, event-based crawls, etc.). The crawler's functional modules will include fetching, extracting, analysing and assessing, scheduling and monitoring, reporting, archiving, and administration. It has been specified that a single, properly configured Pentium-based machine should be able to download 15 million documents per day (roughly 20Mbps bandwidth averaged over the entire day), and that it should be possible to distribute a crawl over up to 20 machines to achieve a 10-fold performance improvement on that of a single machine. Consortium members are committed to open source development so that any local software development by one or more members can be shared and, hopefully, built upon to achieve such harvesting capacity for the use of others. At the same time, members are discussing whether who crawls, stores, and bears long-term archival responsibility for archived Web content can be rationalized.

Working groups

Initial working groups are being formed to discuss archiving file formats and a technical architecture that will ensure appropriate interoperability among modules.

Future projects

Future tool development projects anticipated include an access module (on which work may begin in the fall of 2003), an indexing tool, and post-processing tool(s). Tool development projects will be covered by separate agreements with one of the participating institutions named as lead.

Moving forward

Proposed legislative changes — Bill C-36

As part of the draft bill that is currently before Parliament to establish LAC, the following clause has been included:

> 8. (2) In exercising the powers referred to in paragraph (1)(*a*)[13] and for the purpose of preservation, the Librarian and Archivist may take, at the times and in the manner that he or she considers appropriate, a representative sample of the documentary material of interest to Canada that is accessible to the public without restriction through the Internet or any similar medium.

Such a clause clearly provides LAC with the right to acquire publicly accessible Web sites, and allows a large-scale, non-negotiated, harvesting approach to be legally undertaken. Its implications for long-term public access are somewhat less clear.

Strengthening our commitment

LAC is in the process of redefining itself as a new institution that builds on its NLC and NAC past. Through discussions about LAC's collecting and preservation mandate in the digital age, staff in many areas have become aware of the need to expand the capture of Canadian Internet content. In the short-term future, we need to build institutional capacity in terms of expertise, technical capacity, organizational structure support, and financial support to make a meaningful contribution to the consortium. In the not-too-distant future, whether with the new crawler or using one of the existing tools, we aim to conduct a test crawl of Canadian Web sites and assess its results. There has been little discussion yet as to appropriate parameters of a Web site collection, and further work is necessary to determine precise resource requirements.

Conclusion

LAC's approach to the archiving of Internet content is still maturing. We anticipate that a multi-faceted approach is likely to be needed for the desired level of coverage combined with providing cost-effective description and access. For example, we may choose to undertake large-scale annual crawls of the "surface Web" (e.g. sites in standard formats that are openly accessible and in the .ca domain), and choose to make those available through separate Google-like search interface (with minimal descriptive metadata). Combined with that, we may continue to build a managed and fully catalogued collection of Canadian online monographs and serials, using the crawler tool simply as a way of automating acquisition processes for specific titles. And last, we might concentrate some additional effort to develop custom and negotiated approaches to acquire a limited number of key Web resources that are password-protected or in databases. The goal is to build a rich, diverse, and representative collection of Canadian Web sites for the use of Canadians in the future.

Endnotes

1. National Library of Canada. *Electronic Publications Pilot Project Final Report* (June 1996). www.nlc-bnc.ca/9/4/index-e.html

2. National Library of Canada. *Electronic Publications Pilot Project: Summary of Final Report* (May 1996). www.nlc-bnc.ca/9/4/p4-201-e

3. National Library of Canada. *Canadiana: The National Bibliography*. www.nlc-bnc.ca/6/14/index-e.html

4. National Library of Canada. *Networked Electronic Publications Policy and Guidelines* (October 1998). www.nlc-bnc.ca/9/8/index-e.html

5. Statistics Canada. *Electronic Publications Pilot 1996/97: Final Report*. Ottawa, ON: Statistics Canada, 1997. www.statcan.ca/english/freepub/11-534-RIE/0009711-534-RIE.pdf

6. National Library of Australia. *Online Australian Publications: Selection Guidelines for Archiving and Preservation by the National Library of Australia*. pandora.nla.gov.au/selectionguidelines.html

7. 68th IFLA General Conference and Council, Glasgow, Scotland, August 18–24, 2002. Session 163. www.ifla.org/IV/ifla68/prog02.htm

8. European Conference on Digital Libraries 2002, 2nd Workshop on Web Archiving. bibnum.bnf.fr/ecdl/2002/index.html

9. Masanes, J. "Continuous Web Archiving." *D-Lib Magazine* 8, 12 (December 2002). www.dlib.org/dlib/december02/masanes/12masanes.html

10. Arvidson, A., K. Persson, and J. Mannerheim. "The Kulturarw^3 Project — The Royal Swedish Web Archiw^{3}e: An Example of 'Complete' Collection of Web Pages." 66th IFLA Council and General Conference, Jerusalem, Israel, 13–18 August 2000. www.ifla.org/IV/ifla66/papers/154-157e.htm

11. Cathro, W., C. Webb, and J. Whiting. "Archiving the Web: The PANDORA Archive at the National Library of Australia." Preserving the Present for the Future Web Archiving Conference, Copenhagen, 18–19 June 2001. www.nla.gov.au/nla/staffpaper/2001/cathro3.html

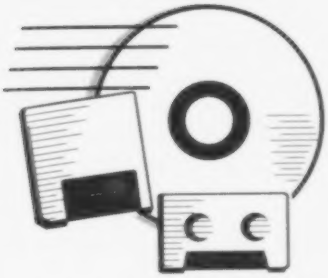
12. Several consortial participants have installed and used existing crawlers such as the NEDLIB harvester or, in the case of the Internet Archive, the Alexa crawler. But all are convinced that a new crawler must be developed. The Internet Archive is in the process of developing its own open source crawler which will replicate the functionality of Alexa. It has not yet been decided whether the consortium crawler project will build upon the new IA open source code, or contract for the complete set of its technical specifications.

13. Paragraph 8(1)(a) reads:
    8. (1) The Librarian and Archivist may do anything that is conducive to the attainment of the objects of the Library and Archives Canada, including (*a*) acquire publications and records or obtain the care, custody or control of them.

Résumé

En 2001-2002, la Bibliothèque nationale du Canada, qui a depuis fusionné avec les Archives nationales pour former Bibliothèque et Archives Canada (BAC), s'est rendue compte qu'elle devait jouer un rôle de premier plan dans l'archivage et l'accessibilité des sites Web importants du Canada. Elle a formulé une stratégie à deux volets pour mieux comprendre les exigences, les mécanismes et les défis associés à l'archivage des sites Web du Canada.

La stratégie était la suivante : 1) lancer un petit projet pilote pour négocier l'acquisition d'un petit nombre de sites canadiens choisis et 2) s'associer à un groupe de bibliothèques nationales et à Internet Archive pour élaborer de concert des instruments et un savoir permettant d'automatiser de grandes collections patrimoniales d'envergure nationale. Nous prendrons acte des progrès accomplis dans le cadre de ces deux initiatives et nous parlerons des prochaines étapes probables.



PRESERVING AUDIOVISUAL RECORDS OF THE UNIVERSITY OF MANITOBA ARCHIVES AND SPECIAL COLLECTIONS

JANE DALLEY

Conservator and Consultant
#407-77 Edmonton Street
Winnipeg MB R3C 4H8
Canada
Tel.: (204) 261-4827
E-mail: JaneDalley@hotmail.com

Abstract

The author undertook a traditional condition survey for the University of Manitoba Archives and Special Collections that became research into commercially available technology and various consumer reformatting/migration options. Unlike the open/shut process of formatting paper-based records to microfilm, digital formats require ongoing monitoring, maintenance, refreshing, and migration. Digital formats were evaluated based on stability, technological obsolescence, ease of migration, and access.

The number of created or born-digital formats in archives is increasing. Is it possible to achieve preservation standards with digital media? Or is the preservation of digital records a systems management issue, requiring an acceptance of digital impermanence and a willingness to commit resources into migration?

Archives must become pro-active in understanding the digital formats that may be acquired as well as those that can be utilized for preservation or access.

"We must learn to manage and migrate collections regularly, and live with impermanence" (Council on Library and Information Resources Risk Assessment Study, 1999).

Part I: The survey

In 2002, the University of Manitoba Archives and Special Collections (UMASC) surveyed its audiovisual holdings in order to identify deteriorating records and evaluate preservation options. The holdings contain only analog media such as film, videotapes, audiocassettes, and reel-to-reel analog audiotape, although UMASC expects to acquire digital audiovisual records in the near future.

Information on the holdings was gathered by externally examining 140 videotapes, 358 films, 1075 reel-to-reel audiotapes, and 1501 audiocassettes. Since the records had been kept in relatively uncontrolled storage conditions, it was decided not to risk playback in case the media had begun to deteriorate. The survey confined itself to physical examination, and testing for vinegar syndrome (a type of chemical deterioration that causes embrittlement and shrinkage of acetate film and audiotape bases). Since film affected by vinegar syndrome cannot be cured and will infect adjacent film, its identification was considered a priority.

Before beginning the survey, the conservator and the Head of the Archives met with the University's Libraries Electronic Technologies and Services staff to develop a spreadsheet that met everyone's requirements. It was decided to incorporate Rules For Archival Description headings, so as to simplify future archival processing. Additional headings were used for storage and physical condition. Descriptors in the headings provided definitions of terms and other associated information for future users. Columns for Use and Value were added for archives staff to set future priorities for reformatting. Although these factors may not be the only ones in deciding priorities, they are a good starting point.

The reel-to-reel audiotapes date from 1955 to 1994 and include a wide variety of brand names and playing lengths. Visual identification of the base and length was required as the manufacturer's box did not always house the original audiotape, and polyester and acetate tape were sometimes spliced together. The physical condition of the audiotapes ranged from very poor to generally good. Many had stepped packs, few had leaders, and even fewer had leaders that were held in place. Loose wind and occasional deformation at the beginning of the tapes were common.

Approximately half of the audiotapes (44%) were on acetate base and many of these tested positive for vinegar syndrome from level 0.5 to level 1.5. Although this is not an advanced level of deterioration, it shows that deterioration is beginning. Older (pre-1970) acetate audiotapes also lose tensile strength as their plasticizers evaporate, and differences between the base and the binder's linear expansion rate can cause tape deformation, edge curling, binder fatigue, and winding tension. The resulting uneven tape-head contact results in poor audio quality.

Most UMASC films were black-and-white prints on both acetate and polyester base. The majority dated from the early 1960s to the early 1980s, with a few from the 1950s and earlier. Many were heavily spliced, with loose leaders or no leaders at all. They were often loosely wound; spoked, curved, and twisted; and deformed at the beginning of the reel. Some films were too big for their reels and coils of film fell off whenever they were moved. Many acetate-based films tested between level 0.5 and level 1.0 for vinegar syndrome.

The audiotapes and films had been stored in a space with few environmental controls and less-than-ideal environmental storage conditions. It was clear that their condition would only get worse, and that the acetate-based audiotapes were a priority for copying or reformatting.

Although the audiocassettes appeared in good condition except for stepped packs, the format is not physically stable. The narrower, thinner tape used in audiocassettes can break, jam, or experience print-through and dropout, especially the 90-minute and 120-minute cassettes that formed the bulk of the holdings.

The UMASC videotapes were also in good physical condition except for stepped packs. They included the U-Matic format, and commercial VHS and Beta formats. The preservation-format U-Matic is unfortunately obsolete. The VHS and Beta tapes are not a preservation format.

The spreadsheet created to identify those records as being in poor condition, unstable, or obsolete now showed that nearly all of the UMASC audiovisual records fell into one or more of these categories. It was obvious that UMASC had to plan for the duplication or reformatting of its audiovisual holdings.

The survey was only just completed when a reel-to-reel audiotape dating from the 1960s broke during playback, and UMASC found itself unable to provide access to an audiovisual record for the very first time.

Part II: Brave new world (and what happened to the old one?)

Until the advent of digital technology, reformatting of audiovisual records usually involved making a new analog master and simultaneously dubbing an access copy or a copy master. From the point of view of stability, fidelity, and ease of migration, analog formats make excellent audiovisual masters. However, few companies still produce the 1/4-inch polyester tape stock and reel-to-reel machines to make and play audiotapes. Although it is still possible to produce a toned black-and-white master film on polyester base (a medium equal to preservation microfilm), the cost is way beyond most budgets.

Why is the stable and standardized analog format obsolete? It is simply because audiovisual equipment and formats are developed primarily for the lucrative consumer market. Consumers are not interested in bulky, low-tech analog formats and the developers are not interested in a low-return market like preservation. The same economic forces causing the demise of analog formats are driving the development of digital technology, and a format's popularity is more important than its stability. Ironically, this is why poor-quality VHS videotapes can still be played, but not preservation-format U-Matics, or film, or reel-to-reel audiotapes.

If UMASC was to continue providing access to its analog audiovisual records, it would need to acquire a range of older audiovisual equipment and maintain it carefully, with the assistance of a technician who understood non-computerized equipment and was able to find the parts. Even if access could be provided to all its existing formats, UMASC would still need to deal with their deterioration issues. Maintaining the analog format appeared to be a dead end.

The University's Imaging and Productions Services was asked for assistance. They offered to copy the deteriorating analog records in-house, using digital technology. Audiotapes would be copied to CD-ROM masters or DAT. Videotapes would be copied to DVCAM digital videotape masters, with DVD format access copies.

A full understanding of this proposed media was required before the offer could be accepted, since reformatting would be the last time that some of the originals were played. Could these or other digital formats re-create the original signal of analog media and provide an archivally acceptable stable master?

Information was gathered from various published sources, commercial suppliers, professional reformatting companies, the Canadian Conservation Institute's workshop *Preservation of Modern Information Carriers*, other university archives, Manitoba Archives staff, and Web sites such as those hosted by the Preservation Research and Testing Division of the Library of Congress, and AV Media Matters listserv@Topica.com.

Suppliers recommended by colleagues were contacted and asked for their suggestions for digital alternatives. Formats recommended for copying the films usually mentioned DVCAM or DVPRO, highly compressed digital formats that have been developed for the consumer market. DAT, a hybrid digital format that puts digital information onto analog tape, was often suggested as a new audiotape format.

It was decided to evaluate these digital media according to the factors of stability, technological lifespan, ease of reformatting and migration, and ease of access.

Stability

The poor-quality manufacturing of optical (compact) discs has improved since the music discs of the 1980s, and they are now thought to last between 5 and 200 years. This rather wide lifespan is narrowed down by other factors such as quality of materials and manufacturing processes, and post-manufacturing factors of environment and handling. Some brands, especially those with gold in their reflective layer, have proved very reliable. Unlike analog records, there is no contact between the reader and an optical disc surface during play, and an optical disc can be played repeatedly without wearing out. If the top metal reflective layer of the disc is lightly scratched, an error detection/correction system can correct lost or damaged data although deep scratches will affect readability.

One of the marvels of digital technology is its ability to store large amounts of information in a very small area. This is achieved by removing redundancies in data and compressing it. Lossless compression allows the data to be compressed as an exact match of its uncompressed state. In order to increase data density, some formats minimize the representation of data by reducing the track spacing, pit length, laser spot size, sampling frequency, and the number of binary elements. This is known as lossy compression. Another method of increasing data density is to use both sides of a disc, or place the information closer together. Compact discs have data on only one side whereas a DVD may have one or two data layers on each side. The DVCAM and DVPRO formats recommended for UMASC use are the brand names of two popular types of recordable DVDs.

DAT is a standardized and commercially supported format. Although it has a built-in error correction capability to compensate for random errors, it has a base tape thickness equivalent to a C-90 audiocassette and is physically fragile. There is a greater chance of damaging the tape during recording and the physical stresses of playback because the head is rotary rather than stationary. The tape can also be damaged through dust. And because the information on the tape is compressed, any damage to or imperfection in the tape affects a lot of information.

Technological lifespan

The issues of machine obsolescence and equipment availability are important factors. In fact, according to the RLG publication *Risk Management of Digital Info*, the most pressing problem in the preservation of digital collections is format and software obsolescence, followed by insufficient resources and lack of planning. Because digital hardware and software are always developing in response to the consumer market, any given format and the technology that supports it will probably be obsolete before the format itself deteriorates. Every one of the digital formats suggested to UMASC for reformatting will be superseded in a short period of time.

Some specialists recommended requesting that any software programs used in reformatting have open (and not proprietary) standards, to increase the chances of playing the program on a number of different systems, and in having access to the technical information necessary for continued access.

Reformatting and migration

Reformatting is the copying of information from one storage medium to a different storage medium, such as when an analog record is reformatted to a digital one, changing it from a continuous signal to a series of binary ones and zeros. Migration is the process of transferring the original sequence of digital structural and data elements to a new configuration or later generation of the same format. Both are major strategies used to address digital obsolescence, and the ease in which this can be done has resource and sanity implications.

All of the digital formats suggested to UMASC would eventually have to be migrated. DAT has an advantage in that it can be copied to a chosen digital format without having to be converted digitally. DVCAM and DVPRO would require migration to the next consumer version, at whatever cost was set by the marketplace. UMASC could find itself on the business end of a game of digital crack-the-whip.

Even with a new format and the requisite resources in place, there is no guarantee that reformatting or migration will go well, as the process of reformatting or migrating can itself generate errors. Different structural elements in the old and new formats can cause internal relationships or data context to be lost, even though content is preserved. Data must be checked for conversion errors after reformatting or migration. Unlike its commercial cousins, an archival record in digital format requires digital files for both its content and its metadata to reduce the chances of this happening.

Metadata is the box, or envelope, in which a digital record sits. Hundreds of metadata elements are required to build the box for each digital record, to ensure that relevant information is saved. The role of metadata is to ensure access and to enable future migration with a minimum of loss. Administrative metadata contain info on how a record has been preserved, what hardware was used, the hardware settings, signal processing, data compression rates, and copyright and other restrictions; structural metadata show the file organization; and descriptive metadata contain info on graphics and text.

A metadata format is laborious to construct but there are pre-made heritage formats, such as the Dublin Core Metadata Initiative, created primarily for textual archives. The Library of Congress is developing a core set of metadata elements for digitized records but has not yet completed the project. The Digital Library Federation has developed the Metadata Encoding and Transmission Standard project (METS), and the Society of Motion Picture and Television Engineers (SMPTE) is reportedly working on a universal format (Metadata Dictionary) that inputs metadata for digital video records at the time of their creation.

Access

Access is one of the basic functions of an archives and must be considered carefully. Access to any format, digital or analog, depends on the availability of the hardware to play it.

Although they may have no intended preservation value, digital formats such as DVD and CD can make good access copies due to their consumer popularity. Affordable and available digital equipment such as CD and DVD players and burners have become standard components in new computer systems. However, there can be a problem with interchangeability between the three common types of CDs. CD-ROMs (read-only) and CD-WORMs (write-once) are standardized and will play on equipment from different manufacturers, but a standard CD player cannot read a CD-RW disc unless it supports the MultiRead specification. A CD-RW drive can read and/or write both CD-R and CD-RW discs, and read CD-ROM discs.

Although analog VHS videotape has been superseded by the DVD format, its continued popularity means that it is still useful as an access format as long as equipment is available to play it.

DAT players can come with various options to enhance playback, and DAT can be numbered anywhere on the tape to permit quick access to a chosen site. However, some playback machines will not play a DAT that was created on another unit. DAT cassettes are small and can be easily stolen.

Part III: Where to, how, and when?

Both analog and digital formats are susceptible to damage from mishandling, poor processing, chemical reactions, and inherent vice. Analog formats are generally considered obsolete and the equipment to play them is disappearing, but digital formats are no different. Digital media are commercially developed formats and the constant change and competition of the consumer market can quickly render them obsolete. The high rate of compression of most digital formats increases the enormity of loss should the files become inaccessible for any reason.

Resources are required to reformat analog records into a digital format, and to maintain a migration program. UMASC felt that it did not yet have the resources to successfully undertake a digital reformatting program. In preparation for an eventual reformatting program, it has begun discussions with the Libraries Administrative Group, the Libraries Electronic Technologies and Services, and Imaging and Production Services with the idea of coordinating and centralizing storage, addressing maintenance and migration issues, and developing joint policies for digital preservation activities. By working together, it is hoped that they can better utilize resources, standardize their policies and practices, more effectively lobby the administration, and achieve economies of scale.

UMASC will monitor developments in digital formats while addressing the immediate needs of the most deteriorated records identified in the survey. Films have been copied to Betacam SP video format, an uncompressed analog format of proven stability. Access copies have been made in VHS video and DVD formats. Although Betacam SP is obsolete, the manufacturer indicates that it will be supported for up to 10 years and UMASC will have advanced its reformatting program before this time.

UMASC considered two options for audiotapes, depending on cost and support. Although making new analog reel-to-reel masters is still considered a viable preservation format by many, UMASC decided to reformat the audiotape as a data file and transfer it uncompressed to a dedicated hard drive or server. By saving it as a data file, there would be an extra level of error correction. The master copy will be copied "as is" without cleanup. For access, the audio file can be burnt to CD. The original tape may be segmented before copying, to facilitate access. All files will be backed up, off-site.

UMASC is keeping original records once reformatted, in case the new format proves unstable or the supporting technology becomes obsolete. However, in the case of a badly deteriorated original record, the digital format becomes the new master.

As well as addressing the immediate needs of its deteriorated records, UMASC will begin to address the larger preservation issue of digital records being acquired or created within the university. No matter the format, digital records must be managed and monitored from their point of acquisition or creation. This can be managed by ensuring that statements on acquisition, conversion, storage, refreshing, migration, and risk management are included in the relevant policies. UMASC will:

- specify which digital audiovisual formats it can and cannot accept to avoid acquiring many different and possibly inaccessible formats
- maintain information on any acquired digital formats so that access and future migration can occur with a minimum of loss
- amend the existing UMASC acquisition policy to address the conversion, storage, refreshing, and/or migration requirements of digital records
- request open standards when purchasing digital software programs or systems to obtain the technical information necessary for managing them
- specify the digital format for an archival master when contracting outside filming of university events

In addition, UMASC plans to keep U-Matic, VHS, and digital playback equipment on site, maintained by trained technicians in the best possible condition. Staff will be trained in "best practices" for the preservation and use of digital formats.

UMASC identified larger preservation issues while investigating the preservation potential of electronic records. It has begun planning for a new archives facility that will provide cool storage for its existing holdings as well as any incoming digital and copy masters. It hopes to include a purpose-built cool/cold vault for storage of acetate-base records. In the meantime, a chest freezer will be obtained for the storage of deteriorated film and reel-to-reel audiotapes.

Conclusion

Many archives are taking on digital preservation responsibilities, either through acquisition or conversion, and yet a 1998–1999 study by the American-based Research Library Group found that 33 of 36 member institutions did not have the policies and practices necessary to preserve and manage digital records (acquisition, conversion, storage, refreshing, migration). In addition, the majority of institutions maintained digital records in at least six different formats. A third of those surveyed had more than 10 formats and two-thirds had digital records that they could not access.

In order to avoid falling into any of these categories, UMASC decided that the responsibility for preserving digital audiovisual records should be the responsibility of all shareholders, from creator to user to caretaker to host. No one group has all the answers but all have relevant knowledge. Together, they must either accept the physical risk of traditional media or commit to the constant change and challenge of digital formats.

Bibliography

Arms, C. *Keeping Memory Alive: Practices for Preserving Digital Content and the National Digital Library Program of the Library of Congress.* Washington, DC: Library of Congress, 2000.

Gibson, G.D. *Bibliography on the Preservation of Magnetic Media.* Preservation Research and Testing Series #9. Washington, DC: Library of Congress Preservation Directorate, 2002, 155 pp.

Gordon, P. (editor). *The Book of Film Care.* Kodak Publication #H-23. Rochester, NY: Kodak, 1983, 138 pp.

Hedstrom, M., and S. Montgomery. *Digital Preservation Needs and Requirements in RLG Member Institutions*. Research Libraries Group, 1998, 43 pp.

Library of Congress. *Risk Analysis Study for a Representative Magnetic Tape Collection*. Preservation Research and Testing Series #9808. Washington, DC: Library of Congress Preservation Directorate, 1998, 20 pp.

Library of Congress. *Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving; the National Digital Information Infrastructure and Preservation Program*. Washington, DC: Library of Congress, April 2002.

Nugent, W.R. *Digitizing Library Collections for Preservation and Archiving*. Preservation Research and Testing Series #9705. Washington, DC: Library of Congress Preservation Directorate, 1997, 58 pp.

Owens, K. *Managing Audio-visual Records in the Government of Canada*. Ottawa, Canada: National Archives of Canada, 1996, 26 pp.

Reilly, J.M. *IPI Storage Guide for Acetate Film*. Rochester, NY: Image Permanence Institute, 1996, 24 pp.

Reilly, J.M. *Storage Guide for Colour Photographic Materials*. Rochester, NY: Image Permanence Institute and the University of the State of New York, 1998, 48 pp.

Reilly, J., D. Nishimura, and D. Burge. *Condition Survey of Motion Picture Holdings in the Library of Congress: Evaluation of Storage Environments for Motion Picture Collections*. Preservation Research and Testing Series #9807. Washington, DC: Library of Congress Preservation Directorate, 1996, 136 pp.

St-Laurent, G. *The Care and Handling of Recorded Sound Materials*. Ottawa, ON: Commission on Preservation and Access, 1996, 18 pp.

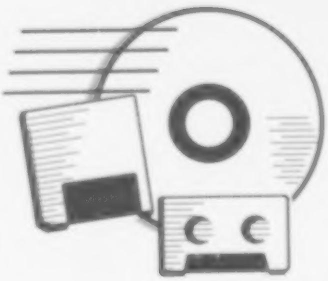
Storm, W.D. *Unified Strategy for the Preservation of Audio and Video Materials*. Preservation Research and Testing Series #9806. Washington, DC: Library of Congress Preservation Directorate, 1997, 78 pp.

Résumé

L'auteur a entrepris un constat d'état ordinaire pour les Archives et collections spéciales de l'Université du Manitoba qui s'est transformé en une recherche sur les technologies disponibles sur le marché et sur l'élaboration de diverses solutions de reformatage ou de migration pour les consommateurs. Contrairement à la procédure relativement simple de préservation de documents imprimés sur microfilms, la numérisation suppose des mesures permanentes de suivi, d'entretien, de recyclage et de migration. Les formats numériques ont été évalués en fonction de leur stabilité, de leur péremption technologique, de leur facilité de migration et de leur accessibilité.

Le nombre de produits numériques transférés ou originaux augmente aux Archives. Est-il possible d'instaurer des normes de préservation pour les produits numériques? Ou la préservation des documents numériques est-elle une question de gestion des systèmes qui suppose l'acceptation de leur caractère transitoire et la volonté de consacrer des ressources à leur migration?

Plutôt que de s'en tenir à un rôle passif, les archives doivent prendre des mesures proactives pour comprendre les formats numériques qu'il est possible d'acquérir et ceux qui peuvent servir à la préservation ou à l'accessibilité.



PRESERVATION STRATEGIES FOR DIGITAL COLLECTIONS: A CRITICAL REVIEW

FRANZISKA S. FREY

School of Print Media
Rochester Institute of Technology
Lomb Memorial Drive
Rochester NY 14623-5604
USA
Tel.: (716) 475-2712
Fax: (716) 475-7029
E-mail: fsfpph@rit.edu

STEPHEN CHAPMAN

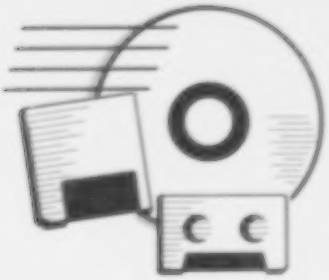
Harvard University Library
Cambridge MA
USA

Abstract

A major challenge in creating digital collections that will survive for a long time is to build systems defined broadly as "digital repositories" that maintain functionality and quality intrinsic to images. It is imperative that the parties involved in creating digital repositories are clear about the difference between "archival" and "deliverable." An archival file has a very low risk factor, meaning that we are confident that neither its integrity nor its functionality will be lost when the format must be migrated in order to remain compatible with image processing applications. A deliverable file can have the same image quality but, depending on file format and compression choices, there is a higher risk of obsolescence — but not total loss if an archival version has also been created and saved. There are various possible scenarios to be looked at when considering the costs of such an endeavor, the main two being payment by file size and payment by traffic/access of the files.

Résumé

L'un des principaux défis à relever lorsqu'il s'agit de créer des collections numériques qui survivront pendant longtemps consiste à bâtir des systèmes définis généralement comme des « dépôts numériques » qui assurent la fonctionnalité et la qualité intrinsèque des images. Il est impératif que les parties qui contribuent à la création des dépôts numériques connaissent la différence entre un « fichier d'archivage » et un « fichier à livrer ». Un fichier d'archivage a un facteur de risque très faible car nous sommes persuadés que ni son intégrité ni sa fonctionnalité ne seront perdues lorsqu'il faudra migrer le format afin de rester compatible avec les applications de traitement d'images. Un fichier à livrer peut avoir la même qualité d'image, mais selon le format du fichier et les choix de compression, il y a un risque plus élevé d'obsolescence; ce ne sera toutefois pas une perte totale si une version d'archivage a également été créée et sauvegardée. Il existe divers scénarios possibles à envisager lorsqu'on considère les coûts d'une telle entreprise, les deux principaux étant le scénario du paiement par taille de fichier et celui du paiement par trafic/accès des fichiers.



COMPUTER-BASED SOUND RECORDING PRESERVATION AND ITS APPLICATION IN SMALL COMMUNITIES

NORM GLOWACH AND JOHN POIRIER

Northwest Territories Archives
PO Box 1320
Yellowknife NT X1A 2P8
Canada
Tel.: (867) 873-7698
E-mail: norm_glowach@gov.nt.ca
john_poirier@gov.nt.ca

Abstract

There are eleven official languages in the Northwest Territories (NWT), nine of which are Aboriginal. The Government of the NWT is committed to the revitalization of these Aboriginal languages, and the NWT Archives is contributing to this endeavour. Since 1979, the NWT Archives has worked at preserving sound recordings, developing a system for preserving sound recordings based on digitizing the recordings in computer file formats. The system includes a database component that documents the work done on the recordings, and also records the type and location of existing versions of the recordings. In cooperation with the Canadian Conservation Institute, the NWT Archives assembled a portable version of the computer sound recording system and transported it to a small community, Rae-Edzo, where people were trained in its use. This paper describes the computer and database systems involved, the project in Fort Rae, and the outcome of the project.

Introduction

This paper describes certain aspects of the sound recording preservation program of the Northwest Territories (NWT) Archives and the first application of this approach, in cooperation with the Canadian Conservation Institute (CCI), in a small northern community setting. As well as providing context for the projects, it provides details of the NWT Archives' in-house preservation program, including current standards. Also included is a discussion of the adaptation of this technology for use in a sound recording preservation project in a small NWT community, Fort Rae.

Context

The NWT Archives is a program of the Culture and Heritage Division of the NWT Department of Education, Culture and Employment. It is located in Yellowknife at the Prince of Wales Northern Heritage Centre, which is equivalent to a combined Provincial Archives and Museum. As well as being the archives of the Government of the NWT, it is the official archives for CBC North and has a legislated mandate to collect non-government records. The Archives has a permanent staff of six, including three archivists, one archival clerk, and two technical services people (the authors of this paper).

The NWT has an area of 1 171 918 square kilometres (somewhat larger than the Province of Quebec), with a population of about 42 000. Of those people, 18 000 live in Yellowknife while the other 24 000 are scattered in 32 smaller communities. Many communities have only seasonal road access, or can be reached only by air. Distances are great and travel is extremely expensive.

Approximately 50% of the citizens of the NWT are Aboriginal. It is unique in Canada in having eleven official languages, nine of those being Aboriginal. As a program of the Government of the NWT, the NWT Archives is obliged to act upon the Government's, and our Department's, stated commitments to the promotion and preservation of Aboriginal languages and cultures. Aboriginal languages have come to the political forefront with a recent review of the NWT Official Languages Act by a Special Committee of the Legislative Assembly. That report places a great deal of emphasis on language revitalization, including preservation. Of late there have also been regular expressions of concern by Aboriginal leaders about the health of their languages.

There is some urgency. Residential schools as well as other forces for assimilation have seriously affected use of Aboriginal languages in the North. Only a few people speak some languages, although others are much healthier. In general, though, the most accomplished speakers are elderly. Transmission of languages to later generations is not occurring at a rate that bodes well for the continued health of the languages. As a result, much of the classical vocabulary and grammar of these languages exist only on sound recordings. Many of these sound recordings are of people who have passed away. Many recordings are held in small communities under poor storage conditions without proper organization or a preservation plan. Important materials have been lost due to arson and poor management.

The number of sound recordings requiring preservation is substantial but not well documented. The NWT Archives holds approximately 9000 items which are well organized. However, it is known through personal communications and visits by Archives staff that thousands of items are held in small communities. The NWT Library System is conducting a survey of all Teaching and Learning Centres (a component of the school system) in the NWT, seeking Aboriginal language materials. As well as printed documents, this survey includes sound recordings. John Poirier is coordinating the development of an Aboriginal languages Web site for the NWT, one of the functions of which will be the exchange of preservation information. These projects should do much to bring materials to light.

The NWT Archives has been aware of the issues of language preservation for many years. This awareness contributed to the establishment of a sound lab working on the preservation both of the Archives' sound holdings and of collections held by other NWT heritage bodies. In addition, the Archives has outreach activities aimed at disseminating preservation awareness and skills to the Aboriginal language communities of the NWT. These activities have been funded in part by the Canadian Council of Archives through the Conservation Plan for Canadian Archival Records program. The NWT Archives also provides safe storage for backup copies of materials held in smaller communities. In several instances it has been able to replace recordings lost or destroyed in their home communities.

Sound recording preservation at the NWT Archives

The NWT Archives was founded in 1979. Our first sound lab, based on reel-to-reel equipment, was established in 1989 and operated by casual/term staff with varying skills until 1994 — at which time the services of Norm Glowach, a professional recording engineer, were acquired on what has evolved into a permanent part-time basis.

In 1993 we developed a long-term work plan for sound recordings based on a collections assessment that has since been updated several times. We first ventured into digital recording (DAT) in 1993, followed by Red Book CD-R recording in 1996 and computer-based recording in 1999. CD-R was chosen as the recording medium because "gold dye" discs have been widely accepted as having good potential for longevity.

The evolution of our recording procedures has been paralleled by the development of an excellent record-keeping system by archivist Ian Moir.

This system not only tracks when and in what format items are copied, but also holds details such as particular items of playback equipment used, CD-R brand names and serial numbers, notes about recording adjustments, and operator names. Appendix A contains screen shots of a portion of the database.

A critical step in the development of our operation came with the 1999 decision to discontinue the process of archiving our audio collections in Red Book Audio format. After careful investigation, the decision was made to switch to computer-based audio file formats stored on CD-R. While Red Book Audio recordings are digital, they are not computer files as such, which has some ramifications where permanent preservation is concerned.

The decision to change to computer-based audio file formats was prompted by questions about the integrity of recordings made in the Red Book Audio format on stand-alone CD-R recorders. In the course of our investigation we learned that Red Book Audio recordings on CD-R potentially have more errors than computer audio files stored on a data CD-R. We also learned that the error correction scheme employed in the playback of a data CD-R has one more layer of recovery than a player for a Red Book Audio CD-R. It became very clear that the data integrity in computer data CD-Rs was superior to Red Book Audio CD-Rs.

An optional process called file verification is used when recording computer audio files to CD-R. As sound recordings are digitized by the analog to digital converter, they are written to a hard drive as audio files. The files are then written to CD-R. After the CD-R is written, the CD-R authoring software compares the file on the CD-R with the original file to ensure that data were transferred accurately, and notifies the user if there is a problem with the data transfer, or the recording media.

When Red Book Audio recordings are made on a stand-alone recorder, the analog signal is converted to a digital stream and recorded to the CD-R without this verification process. It has been found that significant errors can occur that affect both quality and playability of recordings. These errors do not become apparent unless there are audible effects on playback, or the CD-R is checked using a dedicated testing system.

We were able to make our 1999 transition to computer-based recording quickly and inexpensively due to the availability of government surplus Digidesign analog to digital converters, Protools software, and Mac computers. Fortunately our recording engineer was thoroughly familiar with these tools. At that time our recording quality standards (16 bit, 44.1 kHz) were comparable to Red Book Audio.

The validity of our new standard was confirmed in the fall of 1999 in consultations with CCI during a workshop on recording media.

Current approach

Over the last year we have acquired new analog to digital converters, which have allowed us to increase our capture resolution. Our current standards are 24 bit and 44.1 kHz. A complete listing of the equipment and software we now use can be found in Appendix B.

We make two copies of each recording for preservation purposes. Our archival masters currently are in Soundesigner format, the native format for the Protools recording system we use. Backup copies are in 16 bit, 44.1 kHz, WAV format. Archival masters and backups are recorded on two different high-quality brands of CD-R. For reference purposes and eventual Web access we generate MP3 files.

When we started using the 16 bit Digidesign/Protools system, the "Soundesigner" file format was the native format for the system. It would not record in other formats. WAV files were produced by conversion and as a result were a later generation of the recording. We therefore designated the Soundesigner files, being the closest generation to the original recording, as archival masters. When we moved to 24 bit, 44.1 kHz audio files, Soundesigner files were still an option so we continued with this file format for the sake of consistency.

The future

As we look to the future when we go to 24 bit, 96 kHz files, we see that the Soundesigner format will not support those specifications and will no longer be viable for recording archival masters. With this in mind we will re-evaluate which file format will be used for archival masters. We will also monitor the availability of software for playing back Soundesigner files and will migrate as appropriate.

In addition to the data integrity issues discussed above, we see other significant advantages in the use of computer files for sound recording preservation. These lie primarily in the realm of digital asset management. For example, we will have the ability to tie files directly to databases. Given online mass storage we will have the ability to organize, migrate, and make accessible our sound recordings and other media in a very high-quality and cost-effective manner.

Fort Rae pilot project

One very exciting aspect of the evolution of computer-based sound recording is that the cost has gone down very rapidly while the quality has gone up. As a result, excellent equipment and high-quality recording media are far more affordable than was the case for professional analog recording less than a decade ago. This holds much promise for small archival operations.

While comparing notes during the 1999 workshop mentioned above, NWT Archives staff (Norm Glowach and John Poirier) and CCI staff (Tom Strang and Joe Iraci) identified an opportunity for a partnership in disseminating this affordable new technology. CCI had been considering the development of a portable computer workstation for audio recording. The NWT Archives had been thinking along similar lines, had considerable experience with outreach in small communities, and had the technical expertise to train users and maintain a system.

Tom Strang and Norm Glowach (Media Preservation Technician and the NWT Archives' leader on this project) designed and assembled a portable Macintosh system based on Digidesign Protools equipment. It was packed in a custom-built shipping case and delivered to Yellowknife, where a final check and improvements to some components were made (see Appendix C for a listing of system components and software).

While construction took place, Territorial Archivist Richard Valpy made arrangements with the Dogrib Traditional Knowledge Project in the community of Fort Rae (a 1-hour drive north of Yellowknife) to test the equipment. The NWT Archives had previously worked closely with this group in providing advice on the organization and preservation of its work, developing a solid framework for the pilot project.

The system was delivered to Fort Rae in May of 2001, at which time Norm spent several days training staff in its use. The people being trained had elementary computer skills and no experience in sound recording. In some cases their mother tongue was Dogrib. In conjunction with the training, Norm also developed an operating manual for the system. This manual was revised based on feedback during the training process (see Appendix D for an excerpt from the manual).

The preservation project continued to operate until late 2002. Norm made several follow-up visits to reinforce some of the training and to check equipment. The equipment operated very reliably. One staff member developed quite a high comfort level in operating the system. Due to personnel changes in Fort Rae, we do not have a final count of the number of items digitized, although the rate of production seems to have been distinctly lower than our in-house operation. Using sound editing software to check a small number of files, we concluded that quality of work appears to have been very good.

Lessons learned from the Fort Rae project

Our experience with the use of the portable archiving system by the Dogrib Traditional Knowledge Project highlights two points. The first is that it is possible to put together a system and adequately train people having very little prior experience. The second is that the person doing the job must understand the value of the recordings and be committed to doing the transfers well. While it looks exciting at first, the work is actually a very mundane process that requires attention to detail.

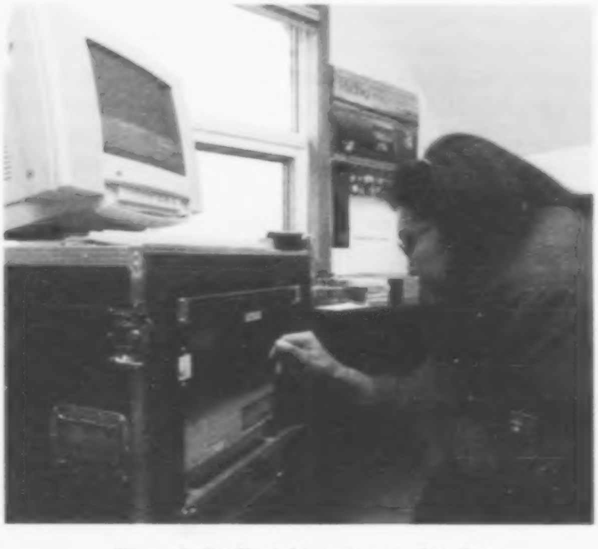


Figure 1. Cecilia Ashton operates the pilot project system in Fort Rae.

Although the trial run at Fort Rae was not a great success in terms of amount of product produced, we were able to show that computerized audio archiving can be done successfully in small northern communities.

Ongoing effects of the Fort Rae project

Since the Fort Rae project, we have heard from a number of other groups interested in sound recording preservation. We have raised the project as a possible model for their efforts. The responses have been varied. Some groups have been reluctant to take on such a project, citing isolation and a lack of local technical support as reasons. Others have been very interested and have purchased their own portable archiving units based on the specs we developed at the NWT Archives. One significant effect has been an increase in requests for training, detailed below.

During the summer of 2003 we helped the Gwich'in Social and Cultural Institute (GSCI) with their audio preservation program. They purchased their own system based on our specifications. We provided training and support.

The most significant component of their system is a CAN$700 Mac/Windows-compatible A/D converter bundled with Protools software made by Digidesign, the MBox.

A GSCI employee, Erin Suliak, worked in the NWT Archives during the summer of 2003. This facilitated training and enabled us to evaluate the MBox for small archives. We were favourably impressed.

As a result of requests from the archival community we have made two major training commitments.

In December 2003 we will run a workshop providing introductory training to seven individuals representing seven Aboriginal languages. Included in that group is an individual from Fort Rae who we hope will provide continuity in that community's efforts.

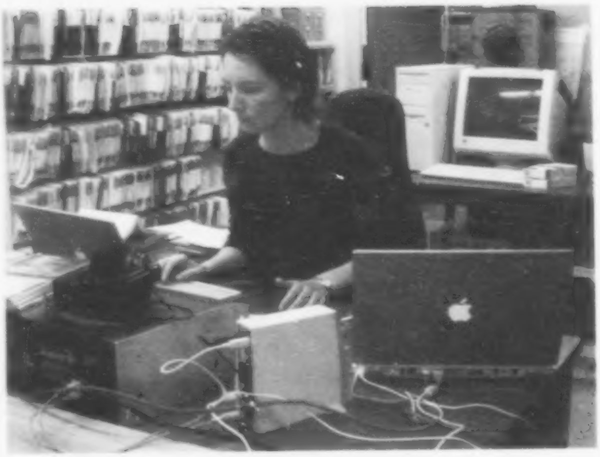


Figure 2. Erin Suliak, a linguist under contract with the Gwich'in Social and Cultural Institute, operates an MBox system in the NWT Archives. The MBox itself is in the centre foreground of the photo.

Selected members of the NWT Archives Council will be given training during Fiscal Year 2004/2005.

Conclusion

We view computer-based sound recording as a valid approach to permanent preservation. It is high-quality, cost-effective, and offers great potential for ongoing management and access in the context of a professional archives.

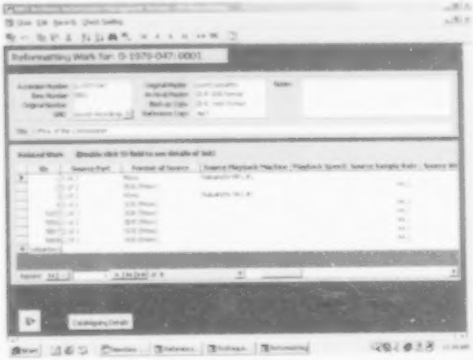
Given the decreasing cost of hardware and software as well as ongoing improvements in user friendliness, we also conclude that under the right conditions it is possible to introduce this technology to smaller archives, even in fairly isolated locations. Those conditions include availability of training with ongoing follow-up from the trainer, support to ensure that reasonable intellectual controls are in place, and adequate local expertise for basic systems maintenance.

Acknowledgements

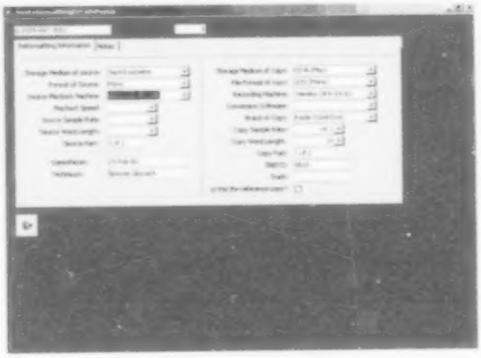
We thank CCI for its support in providing advice regarding sound recording preservation formats and in constructing the portable system used in Fort Rae. We also thank the staff of the Dogrib Traditional Knowledge Project for their cooperation.

Appendix A: Screen shots of selected winows from the NWT Archives database

Reformatting Record Page 1:



Reformatting Record Page 2:



Appendix B: NWT Archives sound digitization recording equipment and software

Mac blue and white 300 MHz G3 computers	Surplus
Digidesign 001 audio interfaces	$950.00
Protools LE software	0.00
Toast 5 CD burning software	$120.00
Norton Utilities 6.0 diagnostic software	$120.00
Techtool Pro 3 diagnostic software	$120.00
Barbabatch 3.1 conversion software	$599.00
Itunes MP3 conversion software	0.00
Quantegy Gold Dye CD-Rs	
Kodak Gold/Silver CD-Rs	

Appendix C: Specifications for CCI/NWT Archives portable sound digitization system

Apple 7200/120 computer
CD-RW drive (internal or external)
Digidesign audiomedia card
Nakamichi MR-1 cassette player/recorder
Furmann AR 1215 regulated power supply
Behringer Powerplay Pro headphone amp
Roxio "Toast" CD burning software
Barbabatch conversion software
Techtool Pro diagnostic software
Rack for all equipment

Appendix D: Sample section of operating manual for portable recording system

Recording Cassettes to Computer (18 steps)

1. Name session and save on hard drive ("Audio Data").

2. Double click on "Audio 1" label in Protools edit screen - Rename the track with an audio file name (Example: Cassette 1A) - Press OK.

3. Click once on "Record" button on the audio track that you have just named - The button will turn red.

4. Place cassette into cassette player - Rewind - Play - Watch audio levels on the audio meters in Protools - Listen to recording in the headphones.

5. Adjust volume on cassette player so that the peaks in audio material (Protools) come as close to top of audio meter scale as possible, without going into the red indicator.

   [The red indicator means that the level is too high and that the sound has clipped. If you set off the red indicator it can be reset by clicking on it once. If the indicator does come on, reset it and turn down the volume of the cassette player and try again.]

6. Once your levels are set, click on the "Record" button on the Protools transport control at the bottom of the screen. It will blink red.

7. Once the cassette has finished recording, press Stop and Rewind.

8. Press the "Return to Beginning" button on the Protools transport, it is located on the far left of the transport, it looks like a line with an arrow pointing left.

9. Press the "Play" button on Protools transport, and the computer will start recording. A red indicator line will start moving from left to right in the Protools edit screen.

10. Press the "Play" button on the cassette player - Listen to recording.

11. Watch red indicator line in Protools edit screen to make sure the sound is not clipping.

12. Press Stop on Protools transport control when side "A" of the cassette is finished - Press Stop on cassette player - Turn cassette over to side "B".

13. Select the hand-shaped grabber tool from the top center of the edit screen in Protools - Click on the audio waveform you just recorded (it will turn black when highlighted) - Push delete on keyboard. Don't worry you are not erasing it!

14. Click once on the two small arrows that point left in the bottom right-hand corner of the Protools edit screen - A box will open on the right side of the Protools edit screen - Check to make sure the name of the audio file you just recorded appears in this box.

15. Return to Step 4 of "Recording Cassettes to Computer" to record side "B" of cassette.

16. Once sides "A" and "B" are recorded, click once on the two small arrows that point left in the bottom right-hand corner of the Protools edit screen - Double click on the audio files you just recorded in the box that appears to the right of the Protools edit screen - The name of the audio file will appear with a number at the end of it (Example: Cassette 1A-01-00) - Double click on the audio file name and erase the –01-00 portion of the name - Press OK.

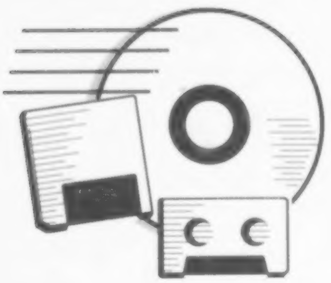
17. File menu - Save session.

18. File menu - Close session.

You have now transferred your first audio recording. To start your next session, begin by double clicking the icon on the desktop "Recorder Template Mono". Then go back to Step 3 and continue.

Résumé

Les Territoires du Nord-Ouest (T.N.-O.) comptent onze langues officielles, dont neuf sont autochtones. Le gouvernement des T.N.-O. est déterminé à revitaliser ces langues autochtones et les Archives des T.N.-O. participent au projet. Depuis 1979, les Archives des T.N.-O. ont travaillé à la préservation d'enregistrements sonores; elles ont conçu un système de préservation dans le cadre duquel les enregistrements sonores sont numérisés pour être conservés sous forme de fichiers informatiques. Le système comprend une base de données qui fournit des renseignements sur les travaux effectués sur les enregistrements et mentionne aussi le type de versions des enregistrements et le lieu où elles se trouvent. De concert avec l'Institut canadien de conservation, les Archives des T.N.-O. ont constitué une version portative du système d'enregistrement sonore informatisé et l'ont transporté dans une petite collectivité, Rae-Edzo, où l'on a montré à la population comment l'utiliser. Cette communication présente le système informatique et la base de données utilisés, le projet qui s'est déroulé à Fort Rae et ses résultats.



PRESERVATION AND THE PRINCIPLES OF ACCESS

ALEXANDER EYKELHOF

Director, Information Technology and Colleges Digital Library
The Bibliocentre - A Division of Centennial College
31 Scarsdale Road
Toronto ON M3B 2R2
Canada
Tel.: (647) 722-9304 or (416) 804-6789 or 1-800-268-5560
Fax: (647) 722-9314
E-mail: alex.eykelhof@bibliocentre.ca

Abstract

The Bibliocentre views the preservation of moving image data as a process of digitization informed by questions of access. Using XML-based metadata description tools, the digitized content is indexed and the user can quickly and easily drill down to very specific moments in the moving image file. The Bibliocentre has worked with TVOntario to develop a concept called the learning moment. This is a segment of the media that stands alone as an educational piece — a learning object. By setting standards for the physical and virtual elements of the content, it has been possible to take boxes of stored tapes out of the vault and deliver their content via the Internet to audiences that would otherwise have had only the most limited access to their knowledge.

Introduction

Digitized preservation of information is a new but growing field and now represents a layer on top of other layers of information. In particular, video content is now being prepared for preservation. Video is itself a relatively new medium and builds on the earlier celluloid formats for moving images. For example, it is possible to view a celluloid copy of the film *Neighbours* by Norman McLaren — as many of us may have done in the 1960s and 1970s. But as time and technology progressed, the film became available to a much wider audience when it was released to video. And now it is available as digitized media in the form of a DVD along with other National Film Board treasures.

Access to stored moving image information is impacted by three factors: the machinery necessary to view the images, the tools available to embed and link pointers to the content, and the availability of the media that actually contains the content. By comparison, archival documents and photographs require little more than a pair of eyes to view them but they are not always made available to the public and are not always accessible in their original form. In many cases the user must be content to work with photocopies or computer-generated copies. But even these still require a set of pointers to make them accessible. A variety of tools are available for that purpose today, not the least of which are the XML-based metadata standards Encoded Archival Description (EAD) and Text Encoding for Interchange (TEI). In many respects these standards have allowed preservation services to provide access to textual or image objects in a manner comparable to the one described in this paper for moving image data. An excellent example of these standards being deployed to provide access to historical papers and photographs can be found at the University of Toronto's Barren Lands Project (digital.library.utoronto.ca/Tyrrell) — a Web site that documents the explorations of J.B. Tyrell in Northern Canada at the beginning of the 20th century.

Perhaps the oldest form of moving image media is celluloid film. Its life spans more than 100 years, carrying records of events from the last century into the present. Many of these events are part of our history and as such we know them intimately, but the recorded sequences themselves have such physical access limitations that it is all but impossible to view the original records. The celluloid is fragile and the machinery is difficult to use; for the most part both machine and film live in museums in climate-controlled environments. What we know of these records is based on their conversion to analog videotape.

Analog videotape has been with us in one form or another for some 50 years now. It is a proven reliable format suitable for private use and for public records but the best predictions for its life range from

10 to about 60 years.[1] Very little comprehensive research has been done on this subject largely because of the newness of the medium. A tape is made up of several components; the polyester substrate layer has a life expectancy of more than 100 years but the binder components deteriorate much more quickly. That being said, some of the earliest tapes dating back to the early 1940s still exist today and are quite playable so it is conceivable that tape life may be underestimated. Also, techniques such as baking or drying the tape are known to extend the life of deteriorating videotapes by arresting the process known as hydrolysis where moisture forms within the layers of the tape.

However, some direction exists on the archiving of the videotape library. Organizations as diverse as the (United States) National Bureau of Standards and the American Film Institute have addressed issues of tape preservation. In a recent survey of digital preservation projects, Beagrie notes "Because the media and technologies used in [the audiovisual preservation community's] industries have been impermanent and cannot be preserved long-term, digitization is widely accepted as their preferred method of audiovisual preservation."[2]

Format issues also play a role in preservation. Having preserved or extended the life of a tape is of limited value if there are no machines available to play it. Hence video content preservation must encompass equipment as well as media. We can be reasonably certain that the formats VHS and S-VHS and even the various Sony Beta formats will be around for some time to come, but the U-Matic format is becoming rare and some of the earlier commercial and professional formats such as the BBC's (British Broadcasting Corporation) Vision Electronic Recording Apparatus (VERA) and the Sanyo VCord system can be considered extinct.

Even if it is possible to retain the equipment in good working order and the content in playable condition, there still remains the challenge of accessing the content or more precisely any specific subset of the content. The machine may play the tape and provide the researcher with a viewable image but short of viewing the entire tape program it will not be possible to identify a specific sequence or image within the tape. The nature of tape is such that access to information is linear and dependent upon some mechanism that counts the minutes/seconds of viewed material. The viewer may know with some degree of certainty at what point a relevant sequence occurs on a tape. He or she may have notes that identify segments of a video and their locations in the time line but will need to view the tape until that point is reached or, if possible, fast-forward the tape to that position. The machinery cannot support direct or random access to the desired points. Hence, access points are available only as adjuncts to the content and cannot be embedded in the content.

The Bibliocentre developed a two-fold approach to address this concern. First, it is advantageous to divorce the content from the format. That is, transform the information on the tape into a medium where options other than linear viewing are possible. Digitizing the content accomplishes this goal. The tape or film media are untouched by this method other than the initial transformation. They may be preserved and protected as a separate and independent process from the preservation of the contents. This is not far removed from Lee et al.'s concept of "Information Migration" which focuses on the intellectual content of a piece and not on its delivery technology.[3]

Second, it is necessary to establish a set of principles by which the availability of the content can be defined. These principles inform the accessibility, providing the developers with an approach that serves the users of archival material without compromising the integrity of that material. Other video projects have cited some of the principles discussed below but not to the same extent. For example, the Internet Moving Images Archive addresses a large amount of video data preserved in digital form but currently provides no detailed access to the contents of the digitized collection.[4]

The project

TVOntario, a publicly funded corporation in the province of Ontario and a leading provider of educational video resources, has a large collection of archived educational material that it had developed initially for television in the early to mid 1980s. This collection includes live action as well as leading-edge animation and has been sitting in TVOntario's vaults for most of the last 10 years in the form of Betacam SP master prints. While this is not a long time in terms of preservation, the corporation needed a solution that would allow it to protect its original investment but at the same time get content back in circulation in the form of an educational service in the kindergarten to grade 12 (K-12) market.

At the time that TVOntario began its search for an organization to assist in these efforts, the Bibliocentre had already begun research and development in this area for wholly other purposes (it was working on a video-on-demand project with the intent of delivering current, college curriculum video content via the Internet using video streaming). This project contained many of the same requirements as the TVOntario project and so, after a careful analysis of the TVOntario request for proposals, the Bibliocentre responded with a proposal based on the study and groundwork already completed for the video-on-demand project. The response was successful and the Bibliocentre was awarded the TVOntario project. A great deal of discussion followed as the two project teams worked closely together to develop the implementation details of the project.

Principles

A set of principles drafted for the project by TVOntario and the Bibliocentre follows. These guidelines were developed from the work of others but also resulted from practical experience at the Bibliocentre and the needs of TVOntario. Some were established in the initial stages of the project while others came to light only after the project had begun.

Identify the audience

When we define the users of a particular set or subset of content it becomes possible to establish taxonomies suitable specifically to that audience. TVOntario had determined who the audience would be based on its own research and planning for this project. It was therefore possible for the Bibliocentre to focus specifically on the needs of that audience and properly codify the content descriptors.

Media independence

The requirement for users or researchers to possess certain equipment or equipment available only in certain places (a well-known problem) often leads to unnecessary limitations to access. The best approach to overcome this problem relies on the use of ubiquitous computer systems, with operating system and browser independent delivery. This was a foregone conclusion as both TVOntario and the Bibliocentre had already recognized the merit of an Internet delivery approach. Lee et al. describe this problem as "Technology Preservation" and while there is value in maintaining "museums of obsolete equipment"[5] it does not suit the purposes of either the Bibliocentre or TVOntario.

Establish the access points

Access points will vary depending upon the standards of delivery and taxonomies deployed, but in general will include such pointers as subject matter, time references, location information, and creation information. After much discussion regarding access points, the primary ones selected for this project were:

- Title
- Subject
- Keyword
- Contributor
- Learning Moment
- Audio Content
- Text Content
- Animation
- Grade Levels
- TVOntario Control Number

Standards-based delivery

To ensure the widest possible long-term availability and interoperability of the content, the data should be described using established metadata standards. The Bibliocentre had begun work with the MPEG-7 standard (the Moving Pictures Experts Group standard for Multimedia Content Description) for multimedia content description and the CanCore protocol which is a subset of the IMS standard (Instructional Management Systems, www.imsproject.org/aboutims.cfm). IMS was designed to be a standard to support distance learning. It specifies interoperability among education software and services. While some modification was necessary to meet the requirements of TVOntario, these were deemed to be within the specification for the project and not detrimental to the more global objectives of both organizations.

Rights management

Digital rights are often embedded in the metadata standard but there are also a number of widely accepted digital rights management standards. Rights management tools will fill an important gap in the delivery of online streaming content but due diligence is essential as well to prevent misuse. In the case of this project, TVOntario owns or had previously cleared the rights to all the material used for this project so it was not an impediment to their delivery.

Provide a context

While it may be possible to identify and locate a 4-minute segment on the Canadian National Railway in a Pierre Berton interview, the information immediately preceding and following those 4 minutes can have great significance in establishing the background or determining the aftermath. The user therefore needs a way to serendipitously identify additional relevant information. The Bibliocentre had explored the use of this concept within both the IMS and MPEG-7 metadata standards and adapted that research to this project.

Support customization
Customization can have two meanings. In the first instance, the user should be able to locally control the view, the screen size, the colours, etc. This level of customization is available in most browsers. At the server level, a "My Archive" feature can support book marking and repeatable queries. The Bibliocentre had considered a "My Library" feature for a later phase of its own video-on-demand project; TVOntario had a number of Web and database designers working on the implementation and considered it as well but also as a later phase of the design.

Support resource discovery
A policy decision needs to be made regarding the extent to which an organization chooses to present its resources over the Internet. Some content may be cleared for generalized searching while other video information may need to be secured and made available only after authentication protocols have been fulfilled. In some cases the metadata can be made available even if the content itself cannot be. This principle is closely related to the initial principle of knowing the audience because it is this group primarily that needs to find the resources.

Support feedback
The value of a resource often depends upon the satisfaction experienced by its users. Because the value of the service is partly a function of its relevance to its users, they need an opportunity to offer insights and provide assistance. This proved to be a boon for the project as it became clear partway through that a new set of deliverables was needed. After discussion between the Bibliocentre and TVOntario, the criteria for the new deliverables were agreed upon and they were produced. These would not have been discovered, however, if TVOntario had not encouraged feedback from its audience.

Metadata schema

The metadata schema for this project was designed with two purposes in mind. First, the Bibliocentre wanted to establish a framework for the discovery of information about the content of the video file. This was done by using a subset of the IMS metadata standard, in particular the CanCore subset, to provide the "envelope information" (i.e. information of a general nature, including title, date of publication, publisher, and a general description) about the video. CanCore was selected instead of Dublin Core (which is also widely[6] used) because IMS is focused on resource discovery in education and therefore more germane to this project. Other work by Jane Hunter describes how it is possible to use multiple schemas including IMS and MPEG-7 within an Application Profiles structure.[7]

The second goal was to provide access to detailed information about the content itself. IMS is not ideal for that and the MPEG-7 schema was selected instead. Its primary purpose is to describe multimedia content to a considerable level of detail. The access points include the subject matter of the content in a particular scene, the spoken content in a scene, the onscreen textual references, the background, and many others.

MPEG-7
Our use of MPEG-7 barely scratches the surface of what is possible.[8] The standard supports access by sound, colour descriptors, sound descriptors, and shape descriptors to name only a few. We are working primarily with the content and are using "description schemes" and "descriptors" to identify the features related to the subject matter. We are *not* interested in (for example) the features shape or motion but are very interested in indexing such features as annotation, actions, and time stamps for those actions.

The primary deliverable for this project is the "learning moment" — a video object extracted virtually (and ultimately physically as well) from an education video. It is a segment typically 2.5–5 minutes in length that stands alone as an instructional segment of video. That is, in those several minutes a learner is able to glean enough information about a subject that we may say he or she has learned something in that moment. For example, in a video on geometry there might be 3 minutes explaining and illustrating the Pythagorean Theorem. Those 3 minutes can be watched in isolation from the rest of the video yet the learner will walk away with an understanding of this important mathematical concept.

Delivering the learning moment required an iterative series of metadata tags designed to describe elements of a video. We will describe the CanCore elements further below but at this point we will concentrate on the MPEG-7 elements. In the parlance of the standard these are *descriptors*. They represent features or characteristics of the video segment that are in some way unique. The repeating sequence of elements informing the learning moment is a *description scheme*. Thus a description scheme may have a number of descriptors. It describes the structure and semantics of a relationship between components that includes descriptors as well as other description schemes. In such a case the description scheme contained within a description scheme would be represented by a description scheme field and related to one or more descriptors. An example of this in the learning moment described above would be where, during the description of the Pythagorean Theorem, a practical example is invoked that illustrates its usage. A background tag would become the description scheme and a descriptor such as keyword would become linked to it.

The *description definition language* (DDL) is the tool that encompasses and allows the creation of description schemes and descriptors. It is based on the XML Schema Language and as such uses the XML Schema structural components including namespaces, element declarations, and attribute declarations. The selection of XML Schema Language for the MPEG-7 (DDL) has several advantages including ubiquity and the ready availability of schema management tools. For an overview of DDL please refer to the article by Jane Hunter in *Introduction to MPEG-7: Multimedia Content Description Interface*.[9]

CanCore

As mentioned above, the envelope information was described using CanCore (a detailed description of this protocol is available at www.cancore.ca/). The information itself, however, was walked across from MARC (MAchine Readable Communication format) records using a transformation tool written by the Bibliocentre for that purpose. The logic behind this decision is based on the fact that most publicly available material has, at one time or another, been catalogued. MARC is a standard format for electronic records widely used in libraries around the world. Our cataloguing staff were able to find 90% of the records needed for this project. The remaining 10% were created manually in MARC format and then transformed into CanCore data. A map was created that linked the common fields in both description formats and a simple PERL program was written to bring over the data from MARC. Important elements transformed from the MARC record include the title (MARC field 245 subfields a and b), language of the video (MARC field 008), subject information of the video (MARC field 659 subfield a), creation information if available (MARC author fields including 100, 110, and 710), date of publication (MARC field 260 subfield c), publication information (MARC field 260 subfields a and b), and description of the contents (MARC field 520 subfield a). Any missing information — information either not in the MARC record but needed for the project or data that did not fit well into the CanCore fields but still deemed important — was manually edited when the metadata librarian analysed the video for the addition of learning moments.

Next steps and conclusion

Possible next steps take us in two directions. First, there is growing trend toward sharing of metadata. These metadata registries are useful for managing metadata and metadata schema and addressing such issues as interoperability between different schema.[10] As more and more metadata are created, the likelihood increases that people will be "re-inventing the wheel" and costly research time will be wasted in duplication of effort.

Finally, the Bibliocentre is interested in working through a consortium licensing model similar to HERON in the United Kingdom.[11] This effort is oriented toward copyright clearance and answers specific needs such as sharing costs and improving the copyright process. As more material is brought forward for digitization and streaming, this issue will become more important.

The project was a major success for both the Bibliocentre and TVOntario, and stands as a milestone in collaboration between organizations with different mandates and purposes. Through the use of metadata and digitization technology, TVOntario was able to inject new life into a collection of archived videotapes. The videotapes themselves have been returned to their home in the storage vaults of TVOntario but continue to be accessed in their digital form via the XML-based encoding. Through the process of building and editing the metadata, the Bibliocentre has enhanced its expertise in various forms of metadata and has had the opportunity to develop a number of procedures for transforming and marking up video content. The Bibliocentre will continue to evolve its repertoire of learning object and video asset management skills. The principles of access described in this paper are a vital component not only to learning object management but also to research in moving image archival material. They will continue to guide and inform decisions about preservation of analog content into digital media.

Endnotes

1. Association of Moving Image Archivists. "Estimating Tape Life." 2002. www.amianet.org/11_Information/11g_VidPres/tapelife.html

2. Beagrie, N. *National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity*. Washington, DC: Council on Library and Information Resources and the Library of Congress, 2003, p. 7.

3. Lee, K.-H., O. Slattery, R. Lu, X. Tang, and V. McCrary. "The State of the Art and Practice in Digital Preservation." *Journal of Research of the National Institute of Standards and Technology* 107, 1 (2002), p. 96.

4. Kahle, B., R. Prelinger, and M.E. Jackson. "Public Access to Digital Material." *D-Lib Magazine* 7, 10 (October 2001). www.dlib.org

5. Lee, K.-H., O. Slattery, R. Lu, X. Tang, and V. McCrary. "The State of the Art and Practice in Digital Preservation." *Journal of Research of the National Institute of Standards and Technology* 107, 1 (2002), p. 95.

6. See for example:
   Hunter, J., and L. Armstrong. "A Comparison of Schemas for Video Metadata Representation." archive.dstc.edu.au/RDU/staff/jane-hunter/www8/paper.html
   or
   Heery, R., and M. Patel. "Application Profiles: Mixing and Matching Metadata Schemas." *Ariad* [illegible] (September 2000). www.ariadne.ac.uk/issue25/app-profiles/

7. Hunter, J. "An XML Schema Approach to Application Profiles." 2000. 10 pp. archive.dstc.edu.au/maenad/appln_profiles.html

8. For a more general examination of the MPEG-7 standard see:
   Martinez, J.M., R. Koenen, and F. Pereira. "MPEG-7: The Generic Multimedia Content Description Standard, Part 1." *IEEE Multimedia* 3 (2002), pp. 78–87.
   For more detailed descriptions visit the MPEG-7 Web site (mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm) or refer to:
   Manjunath, B.S., P. Salembier, and T. Sikora. *Introduction to MPEG-7: Multimedia Content Description Interface*. Chichester, UK: 2002, 371 pp.

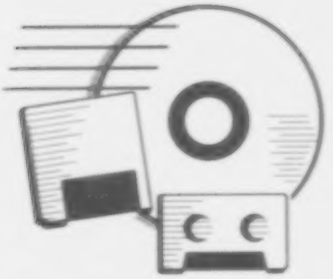
9. Hunter, J., and C. Seyrat. "Description Definition Language." pp. 43–60 in *Introduction to MPEG-7: Multimedia Content Description Interface* (edited by B.S. Manjunath, P. Salembier, and T. Sikora). Chichester, UK: 2002, 371 pp.

10. See for example:
    Blanchi, C., and J. Petrone. "Distributed Interoperable Metadata Registry." *D-Lib Magazine* 7, 12 (December 2001). www.dlib.org
    Another source for metadata registries is from work done in the United Kingdom by UKOLN (desire.ukoln.ac.uk/registry/).

11. Curry, S. "HERON: Project into Service." *D-Lib Magazine* 7, 9 (September 2001), 3 pp. www.dlib.org

Résumé

Le Bibliocentre envisage la préservation de données-images en mouvement comme un processus de numérisation en fonction des questions relatives à l'accès. On se sert des instruments de description de métadonnées XML pour indexer le contenu numérisé de sorte que l'utilisateur puisse se rendre rapidement et facilement à des endroits très précis de l'enregistrement. Le Bibliocentre collabore avec TVOntario pour élaborer le concept de « moment d'apprentissage ». Il s'agit d'un segment du support représentant à lui seul un élément éducatif, un objet d'apprentissage. En appliquant des normes aux éléments physiques et virtuels du contenu, il est possible de retirer de la voûte des boîtes de bandes entreposées et de communiquer leur contenu, par le biais d'Internet, à des utilisateurs qui n'auraient guère eu par ailleurs la possibilité d'en découvrir l'existence.



ARCHIVING ELECTRONIC MEDIA

JIM WHEELER

Tape Restoration & Archival Services
1732 Tierra Nueva
Oceano CA 93445
USA
Tel.: (805) 489-8737
E-mail: Jimwheeler@aol.com

Abstract

Until recently, audio tapes and videotapes have been the main media used for archiving music, voice, radio, and TV. However, this era is coming to an end as digital technology, in the form of optical and magnetic discs, is becoming widely used.

But should archivists switch to digital? And if so, what should they do with their old analog magnetic tapes? If the old data are migrated, finding equipment to play old magnetic recordings will become a problem. Maintaining old equipment is also difficult because of the lack of spare parts and the lack of knowledgeable technicians.

This paper discusses these issues and recommends that archivists network to enter the digital age together.

Introduction

Wire recording was introduced more than 100 years ago, and ever since that time archivists have had to deal with media that were recorded magnetically.

We know that magnetic media can last. Today, there are 100-year-old wire recordings that play on a modified tape recorder with very good sound quality. There are dozens of audio tapes made in Germany in 1943 that can still be played today. That is 60 years! But what about the playback equipment? It is becoming very difficult to obtain video heads for old videotape machines. No one makes spare video heads for obsolete formats, and video heads wear out faster than any other part of a videotape recorder.

Let me begin by briefly covering the main points of preserving the most common electronic media — audio tape (which was introduced to the world in 1948, although it had been invented in Germany in 1935) and videotape (which was introduced by Ampex in 1956 but wasn't used as an archival medium until about 1961).

Magnetic tape

Audio tapes were made with acetate base film until the 1960s, and acetate has the potential to deteriorate unless it is stored in a cool and dry environment. A temperature below 18°C/65°F is recommended, and the ideal relative humidity is 25%. Obtaining such a low humidity usually requires a dryer. Portable dryers (also called dehumidifiers) are available from stores such as Sears. However, I have acetate tapes from the 1940s and 1950s that are still playable today, so we know that some acetate tapes will last a long time in a "normal" household environment.

When videotape was introduced in 1956, it had polyester base film, with Mylar being the most common polyester used. By the early 1960s, polyester was also being used as a base film for audio tapes. In the late 1960s, several of the tape manufacturers started using a polyurethane binder because it had several good properties — including low cost. But in 1978, tests made at the Jet Propulsion Laboratory in Pasadena, California, for the Jupiter Galileo space probe, indicated that tapes with a polyurethane binder had a problem at high humidity.

About the same time, the problem with humidity started showing up in audio tapes and videotapes around the world, and it happened with most of the tape brands. This problem became known as sticky-shed syndrome because of the two different ways in which it could present. In some cases, it would show up as a sticky substance that would stop the tape. In others, it would show up as a powder and would clog the heads. Researchers discovered that this was the result of a breakdown of the binder but that the process could be temporarily reversed by exposing the tape to a dry environment. Eliminating moisture caused the binder molecules to re-link. This drying process made the tape playable for a few days or a few weeks, during which time it could be copied.

However, to play the tape again in the future, it would have to be exposed to a dry environment again. This process could be repeated many times but was not required if the tape was stored in a cool and dry environment.

By the 1980s, some tape manufacturers began to use a more stable binder. But they didn't make a big point of it because that could make them legally liable for past problems. Instead, they used a term such as "more durable." Name-brand tape manufacturers are now aware that the tape they produce is expected to last for many years, but no one will say how many years. One part of the problem is that there is no known test for determining how long a product will last. The other part of the problem is that manufacturers are continually changing their products. When buying new tape, it is best to contact the tape company representative and find out what he/she recommends. All major tape manufacturers have Web sites.

In the 1990s, the manufacturers of videotape recorders stopped making analog machines (except for the consumer VHS VCR) and switched to digital. All of the reasonably priced digital videotape formats use lossy compression, which eliminates fine details like grass, leaves, fine human hair, and the fluffiness of clouds. D5 and D6 formats do not use compression, but these machines cost more than US$60,000 and the tapes are also expensive.

If an original is a poor-quality videotape format, then transferring to a compressed medium such as DVD may not degrade the video. The lowest quality format is the 6-hour VHS. Other low-quality formats include VHS, BetaMax, and 8 mm. Hi-8 and S-VHS are marginal. Originals in any of these formats can be quality tested by transferring them to a compressed digital format and playing the original and the copy side-by-side. To conduct this test properly it is best to use large monitors and view them in a dimly lit room.

Videotapes that are "born digital" should be left in their original digital format.

DVDs

Many people think that DVDs are a good alternative to videotape. But although a DVD can store a feature-length movie, it must be heavily compressed to do so. DVDs are presently capable of storing 4.7 gigabytes (GB) of information, although there is a new DVD standard that will record up to 25 GB on one layer. That is about 12 minutes of uncompressed NTSC video.

CD-Rs and DVD-Rs other than very cheap blank discs have an expected lifetime comparable to magnetic tape. It is recommended that only name-brand discs be used for archival material. For an extremely long life, Mitsui Gold discs (www.mitsuicdr.com) are probably best. Silver is not as good as gold but much better than "normal" blank discs.

Playback equipment

Media that are not exposed to a hot and/or high-humidity environment are likely to last longer than the machines needed to play them. Equipment obsolescence should therefore be a major concern for all archivists. One good example of equipment obsolescence can be seen in desktop computers, which in the past 10 years alone have gone from 5-inch floppy disk to 3½-inch floppy disk to CD-ROM to DVD-ROM. Can you play any of your old 5-inch floppies from 1993? Figure 1 illustrates how storage density has changed over the years.

Maintaining old equipment is very difficult. Not only is there the problem of obtaining spare parts, but a knowledgeable human being and technical manuals are also required. Old technicians have moved, died, or gone fishing.

There are exceptions, and the most notable is open-reel audio tape. Audio tapes made in Germany in 1943 are still playable on machines being produced today. But this will not be true

	Medium	Diameter, inches	Capacity	MB/ sq. inch	
5-inch floppy	Magnetic	5	800 KB	0.04	Audio & Data
3½-inch floppy	Magnetic	3.2	1.4 MB	0.2	Audio & Data
CD-ROM	Optical	4.6	650 MB	50	Audio & Data
DVD-ROM	Optical	4.6	4700 MB	360	Video & Data
DVD-BLU	Optical	4.6	30 000 MB	2300	Video & Data
HDD	Magnetic	3.6	125 000 MB	14 000	Video & Data
HDD	Magnetic	3.6 / two-sided	320 000 MB	36 000	Video & Data

Note: Disc = Optical; Disk = Magnetic

Figure 1. Disk/disc densities over the years.

much longer because digital machines are replacing analog reel-to-reel machines.

The picture should be clear by now. Old collections must be transferred in the next few years or there will be no machines available to make a transfer. Each year, archives should be requesting a budget that includes converting their collection to digital format and making it more accessible.

If at all possible, the original material should be preserved as well as the copy. This may require storing the old collection off-site and in a remote area (cheaper real estate) along with old machines, spare parts, and technical manuals. The storage vault must be climate-controlled.

I recommend that archives within the same region make a survey of who has what tape formats and what types of playback equipment. Armed with this knowledge, these archives can then help each other in maintaining their obsolete equipment and share their knowledge of problems and solutions.

Archivists and librarians in three states have already formed statewide organizations to help with converting collections to digital format, and they each have a Web site:
Colorado: www.cdpheritage.org
Missouri: www.virtuallymissouri.org
North Carolina: ncecho.org

What is meant by "archival"?

When referring to electronic media, the word archival can be used to indicate a number of totally different lifetimes. For many archives (such as the National Archives), the term refers to material that will be kept for at least 50 years and, hopefully, hundreds more. This is the archival media addressed in this paper. On the other hand, for most broadcast television material (with the possible exception of special program material), 10 years is considered a lifetime and many broadcasters use the term archival to describe any material that is not online. For most computer data, 5 years is a lifetime and archival usually describes anything that is a duplicate or a weekly file backup.

Access vs. archival

Compressed videotape formats like DV and DVCPRO make great-looking pictures, the cassettes are small, and the machines have a reasonable price tag. These formats are popular in the broadcast industry, so what is wrong with using them for archiving video? The simple answer is that archivists have a different agenda than television engineers. Archivists must select a medium that has a good chance of surviving for many years, and can be copied without degrading the video or audio. These are not major issues for broadcasters.

For access (viewing) and distribution, S-VHS and VHS are popular formats in many archives, but they are not good archival formats.

Alternatives to videotape and DVDs

Hard disks and datatapes are both good alternatives to videotape and DVDs.

We are all familiar with desktop computers and have been amazed at the continual increase in storage capacity of hard disk drives (HDDs), along with their decrease in cost. When I was shopping for a HDD recently, the smallest drive the store had was 80 GB. And it was only US$120! These two trends (higher capacity and lower cost) will likely continue for several more years.

Last year, I attended the annual HDD trade show (called DISCON) which is held in Silicon Valley each year. About 90% of the desktop HDDs are made by four companies: Seagate, Maxtor, Hitachi (was the IBM HDD group), and Western Digital. At the show, Western Digital announced the release of a drive that stores up to 320 GB and demonstrated a 1 terabyte (TB) drive.

And please note that these are gigabytes and terabytes — NOT megabytes or bits! To put this in perspective:
1 byte = 8 bits = 1 word
1 kilobyte (KB) = 1000 bytes
1 megabyte (MB) = 1000 KB = 1 million bytes
1 gigabyte (GB) = 1000 MB = 1 million KB =
  1 billion bytes
1 terabyte (TB) = 1000 GB = 1 million MB =
  1 billion KB = 1 trillion bytes

Serial ATA is the favourite file format for HDDs. SATA has an interface speed rate of 47 MB per second.

Figure 2 illustrates how the density of magnetic disks has increased on a nearly straight line — a trend that is likely to continue. And with a cost that is now as low as US$1 per GB, the price is right!

Many videos are 2 hours long so archives require a medium that will store 2 hours worth of information.

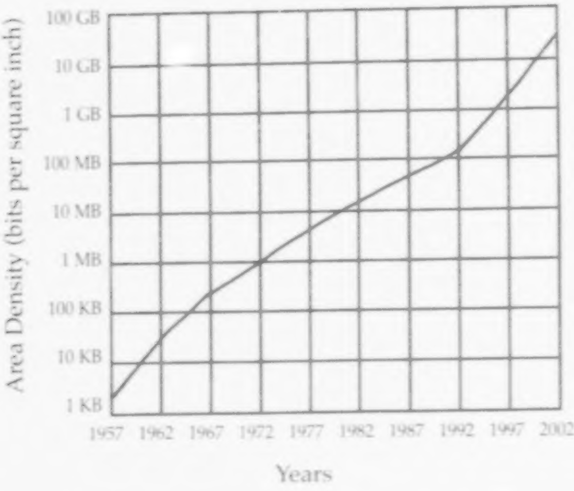


Figure 2. Density of data stored on magnetic disks.

To store 2 hours of uncompressed NTSC video (plus audio) requires 256 GB. The European standards, PAL and SECAM, require around 310 GB. The Western Digital 320 GB drive meets this capacity requirement. At this time, only HDDs can store 2 hours of uncompressed video at a reasonable cost.

Another big advantage of HDDs is that the precision mechanical components are all internal. This eliminates the problem of interchange which is a potential problem with magnetic tapes and optical discs.

HDDs are not the only alternative to videotape; datatapes have also proven to be very reliable and can store as much as 200 GB on one tape. The two most popular formats are S-AIT and LTO. LTO is gaining in popularity and several companies produce equipment for this format.

The costs of these formats are compared in Figure 3. HDDs can be used on just about any computer with an adapter, so there is almost no machine cost for a HDD unless you want to go robotic.

Film

Storing a 2-hour film with no compression requires 5–10 TB. Currently this can be done only by using several HDDs (in a RAID configuration) but a single two-sided magnetic disk will probably be capable of that density within a few years.

Figure 4 illustrates the volume and byte differences of several media.

Format	Machine cost (US$)*	Media cost (US$)
LTO	2200–6500	45–60 (200 GB)
S-AIT	6000–7800	50–60 (25–100 GB)
HDD†		50–400 (80–250 GB)

* The large variation in machine costs is mainly due to the number of slots for tapes that can be in the machine (online) at the same time.

† HDDs can interface with most any computer so machine costs are not significant.

Figure 3. Price comparisons of archival formats.

	1 frame	1 second
35 mm colour film	40 MB*	1000 MB*
NTSC video	1 MB	30 MB

* Cineon system.

Notes:
1 byte (B) = 8 bits = 8 pixels; M = mega = million; G = giga = billion
1 average page of text = about 2500 characters
400 pages of text = 1 MB
1 small DST cassette (VHS size) = 50 000 MB = 20 million pages
CD-ROM = 650 MB
DVD = 4.5–19 GB
HDD = 320 GB
LTO = 200 GB
S-AIT = 100 GB
1 hour of uncompressed NTSC video = 128 GB
1 hour of uncompressed PAL/SECAM video = 155 GB

Figure 4. Comparison of bytes and volume.

A new world

Shifting an archive from videotape to HDDs or datatapes will require a paradigm shift for everyone involved.

First of all, HDDs and datatapes are digital — which makes it possible to send the images electronically. The collection will therefore be more accessible and will likely be used more frequently, so it may be possible to make money from it. No longer need the material sit on the shelves and rarely be used. A computerized digital asset management (DAM) system can also provide quick access to specific subjects and really make the collection user-friendly. However, because DAM systems can be expensive, it is a good idea to ask archival and librarian friends what they recommend before investing in one.

Another major advantage of digital is the ability to keep track of the condition of the recorded material by monitoring the raw (uncorrected) bit error rate

(BER). This should be done with a few samples from the collection about once each year. A radical increase in BER indicates deterioration, and migration is likely required. There is no video quality degradation when migrating uncompressed digital video.

Digitized material also takes less space. Look at the volume required to store 2 hours of video in a non-digital format and compare that with a HDD package that is only 1x4x5 inches (2.5x2.5x12.5 cm).

Although a robotic system is not required, it should be considered as the primary access system as well as a way of automatically transferring material during the night. Robotic systems come in many sizes, and it is possible to start with a small one that can be expanded as funding comes available.

There are several ways of using robotic systems. One that I saw in a TV station has an "On-Line" robotic system that is connected to all of the computers in the station so that everyone has access to the material stored on it. A second robotic system, called the "Near-Line" system, is used for copying the archival material into a compressed form for distribution and "On-Line" use. And the video on the shelves in the archive is referred to as "Far-Line."

Storing HDDs on shelves like books and tapes may not seem "natural" but why not?

It is important to maintain access to old files when software is upgraded (as many have already learned through sad experience with their own desktop computers). My suggestion for coping with the problem of ever-changing software is to dedicate three identical computers to the HDD archival material and NEVER upgrade these computers. The same applies for the electrical connectors and file software that interface the computer with the HDDs. Keep the same old computers and you will always have the same old interface. The reason for the three computers is that, if one fails, it can be used as a source for spare parts to keep the other two operating.

Longevity is a major concern for all archivists, so how long will HDDs last? At this time we do not know. No one knows of any test that will accelerate aging, so the approach is to determine what the potential life-threatening problems are and develop tests to determine if they affect the recovery of the material recorded on the disk.

One way of assuring the longevity of your collection, if there is a disaster, is to make two copies of everything and store one set at a remote location. The most common disasters in tape archives are fire, water pipe breaks, and roof leaks during a storm.

Storage Web sites

www.aittape.com
www.horison.com
www.idema.org
www.sony.com/datasystems
www.storagebysony.com
www.storage.ibm.com
www.ultrium.com

Standards for videotape care

ANSI IT9.23. "Polyester Based Tape — Storage Practices."

SMPTE RP-103. "Care, Storage, Operation, Handling and Shipping of Magnetic Recording Tape for Television."

Conclusion

The "Brave New World of Digital" will require some adjustments and will require all of us to think "outside the box." But it promises to make video collections more accessible and, as a side issue, may help to produce income.

Appendix: A new way to archive video

by Jim Wheeler
www.WheelerTapeForensics.com
October 29, 2003

In the past, video archivists have had to cope with preserving both the media and the machines to play the media. With digital, software becomes a third element.

Another factor is that digital technology changes faster than the previous analog technology. Audio reel-to-reel tape formats have been in use for 50 years. Quad videotape technology lasted for 25 years. With the introduction of computers into the storage equation, we now have to cope with yearly introductions of new operating systems and application programs that may not be able to open old files or play the material in an archive.

Converting video collections to digital is a necessity because none of the old video analog equipment is being produced any more, with the exception of the consumer VHS VCRs. But VHS is not a good archival format.

Other than suicide or retirement, how can an archivist cope with the ever-changing digital technology?

I propose a set of paradigm shifts in the way video material is stored.

First, before we get into the paradigm shifts, think of an archive as a desert oasis — isolated from the rest of the world. The archival oasis includes the storage vault, playback equipment, and computers. Just as caravans can take products to or from a desert oasis, digital archival material can be sent out via electronic wire or fibre optic cable and received via electronic wire or fibre. Archivists are in total control of what happens within the archive. Archivists do not need to do what the rest of the world is doing. The archive is independent. Much like an oasis or island.

In the United States and Canada, NTSC is the standard TV system, so archives must be able to input and output NTSC. NTSC has been around for more than 50 years and will be around for several more decades.

The first paradigm shift is to refrain from upgrading the software on the computer that is dedicated to playing the videos in the archive. This means keeping the old operating system even though new ones are being used by everyone else. NEVER, NEVER, NEVER upgrade the software! Keep the old computers and old software. Because the archive is an isolated oasis, there is no need to care that the rest of the world is constantly updating software.

The second paradigm shift is to have duplicates of all the electronic equipment in the archive. The military and NASA have always required redundant electronics because failure is a certainty. So, how do we make archival electronics redundant? The answer is to start with three identical computers. Three are recommended because that is how many main computers each NASA space shuttle has and because, almost certainly, at least one out of three computers will be usable 20 years from now. At first, only one of the three computers is used. The other two are backups and are not powered.

The first step in setting up a digital oasis is to hire a computer consultant. The computer consultant must select a computer and software that is popular and known to be reliable. In addition, this person must obtain the technical manuals needed to keep these computers operating. This person must be someone who thinks "outside the box" and is not stuck with the idea that the operating system must be upgraded each time there is a new release.

Overheating is a common cause of computer failure. There are at least two ways to counter overheating. One is to post a schedule to clean the computer filters every 6 months (or whatever) and the other is to add a small muffin fan on the back of each computer. This fan will circulate more air over the electronic components and increase their life.

I suggest placing a large banner in the room with the three dedicated archive computers. NEVER, NEVER, NEVER UPGRADE!

In a few years, it is likely that the first computer will have a component failure. When this happens it is time to switch over to one of the other computers and have the first computer repaired without doing any upgrades. A good electronic technician can fix it as long as the technical manuals are available. I have been maintaining radios and tape recorders that are as old as 77 years.

Which media should be used to convert old analog videotapes to digital? First of all, do NOT use a system with compression because compression eliminates some of the fine details in the video. Not using compression means that a high-capacity medium will be required. Presently, the only high-capacity digital medium at a reasonable price is a Hard Disk Drive (HDD).

The third paradigm shift is storing media on shelves in the form of HDDs — which are recommended because each one is self-contained and sealed. One of the advantages of being self-contained is that a complex machine with moving parts is not required to play the video on the disk. It requires only a small adapter add-on to connect with the computer. The major HDD manufacturers have solved the past problems of bearing failure and head crashes. Maxtor, Seagate, Western Digital, and Hitachi (formerly the IBM HDD group) make reliable high-density HDDs.

Ideally, the HDDs should be online in a robotic system, but a couple of external HDD adapters will do as a low-cost starter system. Later, a small robotic (RAID) system can be installed and it can be enlarged in the future.

If the bosses cannot live with the idea of never upgrading the software, there is an alternative:

carry out any upgrades on a second computer and verify that the new software can play the HDDs in the archive. If the new software proves to be compatible with the HDDs, then use the upgraded computer as the main computer and retire the old computer to backup status.

What other choices are available to preserve digital video material? The only other options are to continuously migrate the collection (if it is large) or migrate the entire collection about every 5 years (if it is small). With digital, this is not the daunting task it is with analog — especially if there is a robotic system.

Should a new technology that is an ideal archival medium become available in the future, the archivist can transfer the collection to the new medium using the NTSC (analog TV) video output of your system or CCIR 601 (digital TV).

A major advantage of digital is the ability to keep track of the condition of the recorded material by monitoring the raw (uncorrected) bit error rate (BER). This should be done with a few samples in the collection about once each year. A radical increase in BER indicates deterioration, and migration is likely required. There is no video quality degradation when migrating uncompressed digital video.

In summary, there are four different approaches to archiving legacy video material:

- *Digital oasis* — maintain but do not upgrade the old computers and software until the ideal archival medium is invented
- *Verifiable upgrade* — upgrade software on one computer and verify its compatibility
- *Continual migration* — for a large archive, use a robotic system for continual migration
- *Migration every 5 years* — for a small archive, migrate everything every 5 years

Résumé

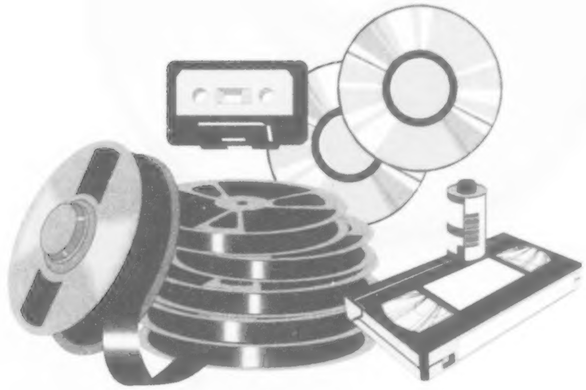
Les bandes sonores et vidéo étaient jusqu'ici les principaux médias employés dans l'archivage d'enregistrements musicaux, vocaux, de radio et de télévision, mais cette ère tire à sa fin. C'est la technologie numérique, sous la forme de disques optiques et magnétiques, qui est aujourd'hui de plus en plus largement employée.

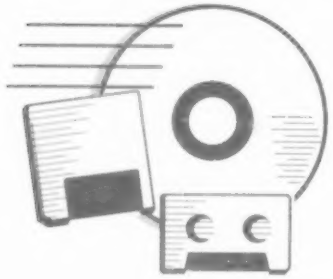
Les archivistes devraient-ils s'initier à la numérisation et que devraient-ils faire de leurs vieilles bandes magnétiques analogiques? S'il faut transférer les vieilles données, il faudra trouver le matériel capable de lire de vieux enregistrements magnétiques. L'entretien du vieux matériel est également difficile en raison du manque de pièces et du manque de techniciens ayant les connaissances requises.

Nous aborderons ces questions et nous recommanderons aux réseaux d'archivistes d'entrer ensemble dans l'ère de la numérisation.

MEDIA KNOWLEDGE

CONNAISSANCE DES SUPPORTS





PRESERVATION OF ELECTRONIC RECORDS: STATUS OF ISO STANDARDS

PETER Z. ADELSTEIN

Image Permanence Institute
Rochester Institute of Technology
70 Lomb Memorial Drive
Rochester NY 14623-5604
USA
Tel.: (585) 475-2303
E-mail: pzapph@rit.edu

Abstract

A subcommittee of the International Standards Organization (ISO) has been writing standards on the preservation of electronic records since 1989. Although the permanence of hardware and software is a major concern, there has not as yet been any standardization work in this area. Instead, activity has been concentrated on the permanence of magnetic tape and optical disc media, and recommended storage conditions to prolong media life have been published for both types of material. Another consumer-oriented standard deals with the care and handling of magnetic tape, and work is underway on a similar document for optical discs. Test methods have been standardized to evaluate the effect of humidity and temperature on optical discs but they do not cover other causes of failure. Another unmet need is a test to determine the chemical stability of legacy materials that are in storage.

This paper reviews the standards that have already been published, those being worked on, and areas where there is no activity.

Introduction

The recent advances in digitizing information have created a technological revolution that is having profound influences on society. The advantages of digitizing information with regard to access, data compaction, manipulation, transmission, and distribution are very clear and very well known. However, the effect of digitizing on information preservation is much less evident and generally not positive. The only positive effect on preservation is that digital images can be accessed — thus eliminating the need for handling originals with the resultant degradation caused by wear and tear. The major problems are those associated with hardware and software considerations and the permanence of the media itself. The international standards that have been published, that are under preparation, or that are not being addressed provide a good overview of the current status of digital image permanence. The status of these international standards is the focus of this paper.

Standards

The International Standards Organization (ISO) is the recognized standards body in the world with responsibility for the permanence of images and information. Its headquarters are in Geneva and its standards are prepared by various technical committees, the technical committee on image permanence being TC42. Membership of TC42 is through the national standards organization of various countries. For example, Canada is represented by the Standards Council of Canada and the United States by the American National Standards Institute (ANSI). Twelve countries are represented on this committee.

There are four types of standards that have been developed in the field of image permanence.

- Definitions — It is important that all who are involved in this field use the same words to mean the same thing. To eliminate possible confusion, representatives from the photographic, magnetic, and optical disc technologies have agreed on several important terms such as the meaning of "life expectancy" and the distinction between "medium-term" and "long-term" storage conditions.

- Specifications — Specifications are minimum property requirements for materials to meet a classification with respect to longevity. A considerable number of specifications have been published for photographic film. Unfortunately none exist for digital images nor are any to be expected in the near future.

- Standard practices — Standard practices are recommended operating procedures. Here there has been significant progress for electronic images. Examples include the conditions for storage of both magnetic tape and optical discs as well as handling recommendations for magnetic tape.

- Test methods — Test methods have been issued as separate documents to standardize the means to evaluate specific properties.

There is an ever-increasing interest in the preservation of recorded information — a result of the passage of time and the unfortunate loss of many valuable and historical records. This concern is reflected in the increase in published standards on photographic and digital image permanence, from only one in 1945 to 23 today. Several more are in preparation. All ISO publications in this field carry the numerical designation ISO 189XX. This uniformity in the numbering system facilitates identification for this field of interest.

Critical components

For human-readable information such as photographic images or printed records, only the permanence of the media itself is critical. But this is not the case for digital information where hardware and software must also be considered. The critical components discussed in this paper therefore include hardware and software, magnetic tape, and optical discs.

Hardware and software

The preservation of the necessary hardware and software is a major concern in the recovery of information on magnetic tape and optical discs. Unlike photographic products, these materials can only be machine-read, and the problem this poses has been recognized for many years.[1,2] As equipment is used, parts wear out and must be replaced. But equipment repair is only possible if components are still manufactured or can be cannibalized from other machines. And if past history is any guide, hardware becomes obsolete within a relatively short time. Systems in this field are constantly undergoing change, and it is very doubtful that replacement parts will be available after several decades. There are many examples where information on tape has been lost or would be very expensive to retrieve.[3]

In the face of this continued evolution and consequent obsolescence of playback hardware and software, consumers might wish that the standards committees would create a single readability (i.e. hardware–software) standard to which all future electronic products would have to adhere. However desirable this approach may be from a storage perspective, it is not at all practical because it would inhibit the development of technology, retard improvements, and never be supported by the manufacturers. To illustrate the impracticality of standardizing on a single format, consider the history of sound recording. Over the last century audio recordings progressed from wax cylinders to 78 rpm records to 33 rpm records to magnetic tape cassettes to CDs to DVDs. Standardizing any of these technologies would have had no practical effect on creating a single readily accessible system. There are thus no plans in any of the standards committees concerned with permanence to standardize hardware and software.

The only practical approaches to preservation are therefore either a well-managed program of reformatting stored data from one system to a newer one as obsolescence proceeds or the use of a human-readable material such as paper or film.

Reformatting digital data presents some practical problems.

- It is difficult to predict when a format becomes obsolete and when the information must be transferred. Failure to do so in time can be a catastrophe.
- Although digital information can be copied without any loss of quality, it is an expensive operation. This expense results both from the copying process and from maintaining the digital archive.[4-6] The National Archives[6] in the United States has estimated that the cost of maintaining a digital archive during the first 10 years is likely to be 50–100% of the original reformatting expense. Reformatting must therefore be considered a yearly recurring cost, and some archives may find this difficult.
- It takes only one inadvertent failure to reformat during the desired life of a collection for the information to be lost forever. For example, if reformatting is required every 20 years, this requires five reformat procedures each century. Natural disasters, wars, or staff neglect are obvious possibilities for failing to reformat on schedule.

The alternative that is proposed at this time is to preserve records in human-readable form. This can be text that is printed on paper or information obtained as photographic images.[7,8] And of course the original human-readable information that is digitized for accessibility, display, manipulation, or transfer must be retained. Unfortunately this does not address the problem of information that is obtained or available only in digital form or on tape or discs.

Magnetic tape

The useful life of hardware and software will certainly place a limitation on the permanence of digital information on magnetic tape. However, it is very desirable that the stability of the tape media not be the limiting factor. Many tape collections in government archives, broadcasting and TV networks, or cultural organizations are much too large to consider periodic reformatting or copying. For these reasons there is a very strong incentive to use tape that has an extended life expectancy and to store and handle it to prolong its life as long as possible. It is with this objective that a task group was organized in 1989 to prepare standards on these topics. Initially this task group included a number of manufacturers, various archives and libraries, and several restoration and research laboratories. Unfortunately there has been a considerably reduced involvement by the tape manufacturers in recent years. This abstention may possibly be due to the general realization that inherently tape does not have a long life expectancy.[9-13] With limited resources it is not realistic to expect the manufacturers to continue studies on an inherently weak property. The withdrawal of the tape manufacturers has severely restricted the laboratory facilities available to the task group for the needed experimental work. Despite this restriction, a number of significant projects have been completed with considerable input from several archives and restoration laboratories.

A standard has been published[14] on storage conditions, and it incorporates recommendations made in several publications.[15-18] This document discusses the importance of tape wind and tape pack integrity, external magnetic fields, recommended storage housings, storage rooms, fire protection, and monitoring of the tape collection. Two types of storage conditions are specified, one intended for medium-term storage (providing a life expectancy of 10 years) and the other for extended-term storage. The extended-term storage specifications call for more rigid environmental conditions that are expected to give a useful life of 50 years.

Degradation of magnetic binders is caused by chemical reactions, the rates of which can be lowered by decreasing the temperature and relative humidity (RH). Consequently, the useful life of magnetic tape can be increased by lowering either the storage temperature or storage humidity. Moreover, a lower storage temperature can compensate for a higher humidity to obtain the 50-year life expectancy.[10,13,19,20] However, it is important that the storage temperature not be lower than 8°C because of concern that at such low temperatures the lubricant may separate from tape binder and tape pack loosening may cause permanent tape distortion. There has been no published information on this possible behaviour, so this is an area where experimental studies are clearly needed.

A more recent standard[21] deals with the care and handling of magnetic tape. Tape is relatively thin (varying from 10 to 50 μm), which allows for greater data compaction but results in greater fragility of the material. Consequently care in handling is very important. This standard gives recommendations on maintaining the tape pack integrity and tape tension, and on procedures for transportation, cleaning, staff training, and inspection. There are also discussions of problems associated with contamination, damage caused by disasters, and other environmental concerns. This document was prepared primarily with input from users who have had practical experience; laboratory studies were not required.

Although these two published standards represent significant guidelines for the magnetic tape consumer, there are still a number of unfulfilled needs. As mentioned above, laboratory work is still required on the advisability of very low temperature storage.

Another need is a specification that can predict the life expectancy of tape. This is a rather formidable task considering how tape is used, i.e. in recorders and players that require it to be transported at high speeds from hubs and reels over small-diameter rollers and magnetic heads in a tremendous variety of configurations. Because tape is inherently fragile and is subjected to this kind of physical abuse during normal use, its life expectancy is usually controlled by physical failure. This physical failure can be manifested in many ways depending upon the format, the recording device, and the playback equipment. One critical feature in magnetic systems is the close spacing between the magnetic head and magnetic medium. This is required for optimum output, but it causes head wear, binder wear, or both. In some situations, the tape fails due to binder debris clogging the magnetic heads; in other situations the failure may be due to high friction and the resulting lack of uniform transport.

A specification for tape life expectancy requires tests to cover all these various types of failure. There is general agreement that such a specification should include a friction test, a binder-support adhesion test, a method to determine the abrasion resistance of the binder, and a means to evaluate the chemical stability of the magnetic layer. The stability of the magnetic output must also be established. In addition, it is very desirable that a magnetic tape specification not be format specific but rather be applicable to all widths and tape types. This eliminates the possibility of using recording or playback equipment to evaluate characteristics such as binder wear or head clogging. Once the test procedures are established, accelerated incubation conditions must be determined that will accurately reflect behaviour of aged tape under operating and storage conditions. To predict the life expectancy under these conditions, incubations must be carried out at a series of elevated temperatures and extrapolations made to room conditions using the Arrhenius treatment[22] of the data. This not only requires the participation of organizations with laboratory facilities, but also involves a considerable expenditure of staff time. The resources to pursue this objective are just not currently available.

The magnetic tape task group recognized that obtaining a life expectancy specification may not be a realistic goal at this time. However, it was suggested that a standard that just rank-ordered tapes with respect to their long-term stability would be of benefit to some consumers. This would eliminate the need for an Arrhenius type study but it would require establishing standardized test methods for the properties previously listed. While this would not require the large expenditures needed for a tape specification, it would still require considerable laboratory participation.

Still another unmet need for the tape archivist is a means to analyse the condition of legacy tapes that are in storage. This would enable the archivist to establish priorities for tape duplication. Such a procedure would have to be non-destructive and relatively quick, otherwise it would not be at all practical to evaluate tape stability in a large collection. Obviously running the tape in playback equipment would not be realistic. At the present time, tapes in storage are subjected only to a visual examination as described in the care and handling standard.[21] This involves inspection for integrity of the tape pack such as popped strands, stepped pack, edge damage, or loose wind, evidence of particulate contamination, odour, or signs of fungus. Unfortunately, this type of testing does not necessarily detect any binder degradation that could cause subsequent problems in playback.

The most ideal technique would be an acid detection indicator since high acidity is a sign of existing or incipient tape degradation. Such an approach, employing A-D Strips, has been used to determine the stability of cellulose triacetate base photographic film.[23,24] (A-D Strips are pH-sensitive papers that detect the presence of volatile acetic acid resulting from the acetate base decomposition.) However, a comparable approach might not be successful for magnetic tape as the acid produced by the binder degradation is not very volatile. It is possible that a surface pH measurement might be practical. The Image Permanence Institute has submitted a grant proposal to fund such a study.

In summary, the standardization process has produced several documents that can be of significant benefit to the magnetic tape user. However, there are unmet needs such as a specification dealing with tape permanence and a quantitative or semi-quantitative means to determine the extent of degradation of tape materials in storage. The absence from standardization work of a critical number of tape manufacturers with their laboratory facilities makes progress very difficult.

Optical discs

The situation with optical discs parallels in many ways that with magnetic tape. Here too hardware and software considerations limit long-term preservation unless the information is reformatted, and it is very desirable that the permanence of the optical media not be the critical factor. An optical disc task group was also organized in 1989 with the purpose of writing standards that can determine and extend the life of the discs themselves. A number of disc manufacturers initially participated in this activity. However, none are involved today. It is believed that in this highly competitive market, the scientific resources of the manufacturers are concentrated in developing new formats that provide an advantage with respect to size or data compaction. Resources are not available to establish stability standards of existing products.

Optical discs have an important advantage over magnetic materials with respect to predicting their life expectancy, i.e. because optical discs are recorded and read by light, they do not come into contact with moving or stationary parts of equipment. Therefore their useful life is mainly determined by the properties of the material itself, and, unlike magnetic tape, physical wear and tear is less of an issue. About a decade ago, several manufacturers used the Arrhenius method to predict the useful life of their products.[25-27] Investigations were also completed by the Library of Congress[28] and the National Institute of Standards and Technology[29,30] in the United States. All these studies used an electronic readout signal to measure the course of disc degradation. As a result of this activity, the optical disc task group wrote three test methods that predict the effects of temperature and humidity on the useful life of CD-ROM,[31] MO,[32] and CD-R[33] discs.

It should be noted that these are test methods to predict the effects of temperature and humidity only and they are not material specifications. The difference is that the latter give test requirements for all aspects of permanence that a material must meet to be given a life expectancy rating, while the former is only a single test procedure. Optical discs can fail by a number of different mechanisms, such as relaxation of the substrate causing warping, changes in the reflecting layer by corrosion, cracking or pinholes, changes in the reflection of any dye layers by light, pressure or crystallization, or breakdown of the disc laminate by adhesion failure and layer separation. The three standardized test methods are quite explicit in stating that they are only valid if the dominant failure mechanism at the accelerated conditions in the test method is the same as during usage. They also do not address changes due to light, exposure to corrosive gases, or rapid changes in temperature or humidity.

Temperature cycling may be critical because the different component layers have different thermal coefficients of expansion, resulting in large stresses between layers with the possibility of cracks, adhesion failure, or deformation. Humidity cycling is likewise important because of differences in dimensional changes due to moisture absorption. Potential problems due to humidity cycling are more important for discs than for magnetic tape because of the greater thickness of the component layers in optical discs, the longer times for moisture equilibrium, and the greater forces involved. The importance of temperature–humidity conditioning times has been recognized[34] in the test methods, and staging times are specified. Environmental cycling tests are useful to determine disc behaviour under high stress conditions; however, they cannot be used for predictions of life.[28,35]

The limitation of these three test methods is that they apply only to failure by temperature and humidity. Moreover, they are complex and time-consuming procedures and must be carried out by an experimental laboratory with specific scientific skills. Their main advantage is that they provide a recognized and sound technical procedure for developers of optical discs to optimize the stability of discs to these variables.

Of particular interest to the consumer is how long optical discs will last. Most of the published information is based on the test methods described above. Some manufacturers[25,26,27] predict life expectancies greater than a century. The National Media Lab in the United States[36] reported a very wide range — from 5 to more than 100 years depending on the product. The absence of a specification for the life expectancy of optical discs represents a significant void in this technology. However, because of the technical complexity involved in preparing a specification and the absence of experimental laboratories that would participate in this activity, optical disc specifications are unlikely. It should be noted that similar experiences have been found with other materials. Although a specification has been written for paper,[37] which is a very mature technology, it is currently undergoing a major review. Likewise, while specifications exist for four types of black-and-white photographic film,[38-41] there are none for colour film or colour paper.

Although progress in publishing standards on optical disc permanence is quite limited, a positive accomplishment has been the standard on optical disc storage.[42] This document incorporates recommendations made in several publications.[36,43] A RH between 20 and 50% and a temperature below 23°C is specified. As with magnetic tape, protection is increased by storing discs at low temperature and low RH — although storage below -10°C and 5% RH is not recommended because of the possibility of damaging the mechanical integrity of the optical disc laminate and disc brittleness. Cycling of RH should also not be greater than ±10% over a 1-day period. Particular care should be given to maintaining a low dust and dirt environment. Protection from light is vital for many writable CDs and magnetic fields are a concern for magneto-optical discs.

A current activity underway by the optical disc task group is the preparation of a standard on the care and handling of optical discs. This is expected to be a companion document to the standard on magnetic tape.[21]

The standardization accomplishments on optical disc stability are definitely limited. They reflect the current status of knowledge in this field which is not a satisfactory situation. However, it must be kept in mind that this technology is less than 20 years old. And when other image recording materials such as paper, photographic film, and magnetic tape were a comparable age, standards did not exist for them either.

Conclusion

The status of digital image permanence is reflected in the ISO standards that have been written and also by the subject areas that have not been addressed. Consumer-oriented documents on storage are available for both magnetic tape and optical discs and a handling standard has been finalized for tape. Test method standards on the effects of temperature and humidity on three types of discs are of importance to a limited group. However, specifications have not been standardized on the permanence of either magnetic tape or optical discs and there is no current activity on such standards.

The permanence of digital information is believed to be satisfactory for most users with relatively short-term or medium-term interests. However, it is possible that for long-term requirements of a century or more, the only currently available way to ensure the preservation of information is in a human-readable form.

Endnotes

1. Mallinson, J.C. "Archiving Human and Machine Readable Records for the Millennia." pp. 388–403 in *Second International Symposium: The Stability of Preservation of Photographic Images — Printing of Transcript Summaries*. Springfield, VA: Society of Photographic Scientists and Engineers, 1985.

2. National Archives Records Administration. *Preservation of Historical Records*. Washington, DC: National Academy Press, 1986.

3. Sniffen, M.J. "Lost History." The Associated Press, APN 01/02/0009, January 2, 1991.

4. *Preserving Digital Information*. Report by the Commission on Preservation and Access and the Research Libraries Group, 1785 Massachusetts Ave. N.W., Suite 313, Washington, DC 20036, USA, August 1995.

5. Lesk, M. *Preservation of New Technology*. Report by the Commission on Preservation and Access, 1785 Massachusetts Ave. N.W., Suite 313, Washington, DC 20036, USA, October 1992.

6. Puglia, S. "Overview of NARA's Electronic Access Project." Presentation at the Symposium: Preserving Photographs in a Digital World, Rochester, NY, August 2001.

7. Wallace, J. "Considerations Regarding the Long-term Storage of Electronic Images." Presentation at the Digital '91 Photography Conference, Dallas, TX, USA, February 1991.

8. Easton, R. "National Archives Approach to Film, Sound, and Videotape Archiving." *Image Technology* 73 (March 1991), pp. 87–93.

9. Bertram, N., and A. Eshel. "Recording Media Archival Attributes (Magnetic)." RADC-TR-80-123. New York, NY: Rome Air Development Center, April 1980.

10. Bertram, H.N., and E.F. Cuddihy. "Kinetics of the Humid Aging of Magnetic Recording Tape." *IEEE Transactions on Magnetics* MAG-18, 5 (September 1982), pp. 993–999.

11. Bradshaw, R.L. "Archival Stability of Flexible Magnetic Media." Presentation at the Annual Conference of the Association of Canadian Archivists, Fredericton, NB, Canada, June 1989.

12. Nilsson, E., and M.L. Samuelsson. *Investigations on Computer Tapes for Information Storage at 6250 BPI*. Bovås, Sweden: Swedish National Testing Institute, 1990.

13. Smith, L.E. "Factors Governing the Long-term Stability of Polyester-based Recording Media." *Restaurator* 12 (1991), pp. 201–218.

14. ISO 18923. "Imaging Materials — Polyester-base Magnetic Tape — Storage Practices."*

15. Jenkinson, B. "Long Term Storage of Video Tape." *British Kinematography, Sound and Television Society Journal* 63, 3 (March 1982), pp. 126–127.

16. Geller, S.B. *Care and Handling of Computer Magnetic Storage Media*. Computer Science and Technology Series, NBS Publication 500-101. Washington, DC: U.S. Government Printing Office, June 1983.

17. Wheeler, J. "Increasing the Life of Your Audio Tape." *Journal of the Audio Engineers Society* 36, 4 (April 1988), pp. 232–234.

18. Van Bogart, J.W. *Magnetic Tape Storage and Handling*. Report by the Commission on Preservation and Access, 1785 Massachusetts Ave. N.W., Suite 313, Washington, DC 20036, USA, June 1995.

19. Brown, D.W., R.E. Lowry, and L.E. Smith. "Equilibrium Acid Concentrations in Hydrolyzed Polyesters and Polyester–Polyurethane Elastomers." *Journal of Applied Polymer Science* 28 (1983), pp. 3779–3792.

20. Van Bogart, J. Presentation at the Symposium: Salvage of Magnetic Media Emergency Preparedness and Response for Libraries, Archives, and Museums, Library of Congress, Washington, DC, USA, April 28, 1995.

21. ISO 18933. "Imaging Materials — Magnetic Tape — Care and Handling."*

22. ISO 18924. "Imaging Materials — Test Method for Arrhenius-type Predictions."*

23. Fischer, M.C., and J.M. Reilly. "Use of Passive Monitors in Film Collections." pp. 11–40 in *Topics in Photographic Preservation: Volume 6*. Washington, DC: American Institute for Conservation of Historic and Artistic Works, 1995.

24. Adelstein, P.Z., J.M. Reilly, D.W. Nishimura, and C.J. Erbland. "Stability of Cellulose Ester Base Photographic Film: Part III — Measurement of Film Degradation." *Journal Society of Motion Picture and Television Engineers* 104, 5 (May 1955), pp. 281–291.

25. Murray, W.P. *Life Expectancy of Optical Systems*. Report to ANSI / AES Joint Technical Commission IT9-5, October 1990.

26. Murray, W.P. "CD-ROM Archivability." *NML Bits* (Newsletter of the National Media Lab, St. Paul, MN, USA), May 1992.

27. Stinson, D., F. Amell, and N. Zaino. *Lifetime of Kodak Writable CD and Photo CD Media*. Digital and Applied Imaging Report, Eastman Kodak Co., 1994.

28. Nugent, W.R. *Estimating the Permanence of Optical Discs by Accelerated Aging and Arrhenius Extrapolation*. Report to ANSI Technical Committee X3B11, June 1989.

29. Podio, F., R. Onyshczak, and E.S. Villagran. "Standardization of Testing Methods for Optical Disc Media Characteristics and Related Activity at NIST." *Optical Information Systems* (July–August 1990), pp. 175–178.

30. Podio, F.L. *Development of a Testing Methodology to Predict Optical Disc Life Expectancy Values*. NIST Special Publication 500-200. Gaithersburg, MD: National Institute of Standards and Technology, December 1991.

31. ISO 18921. "Photography — Compact Discs (CD-ROM) — Method for Estimating the Life Expectancy based on the Effects of Temperature and Relative Humidity."*

32. ISO 18926. "Imaging Materials — Life Expectancy of Information stored on Magneto - optical (MO) Discs — Method for Estimating, based on Effects of Temperature and Relative Humidity."*

33. ISO 18927. "Imaging Materials — Recordable Compact Disc Systems — Method for Estimating the Life Expectancy based on the Effect of Temperature and Relative Humidity."*

34. Wrobel, J.J. "Ramp Profiles for Optical Disk Incubation." pp. 191–202 in *1994 Topical Meeting on Optical Data Storage*. Bellingham, WA: SPIE (International Society for Optical Engineering), 1994.

35. Marchant, A.B. "Optical Recording — A Technical Overview." Boston, MA: Addison-Wesley Publishing Co., 1990, Chapter 14.

36. *NML Bits* (Newsletter of the National Media Lab, St. Paul, MN, USA), March 1995.

37. ISO 9706-1994. "Information and Documentation — Paper for Documents — Requirements for Permanence."*

38. ISO 18901-2002. "Imaging Materials — Processed Silver-gelatin Type Black-and-white Film — Specifications for Stability."*

39. ISO 18905-2002. "Imaging Materials — Ammonia-processed Diazo Photographic Film — Specifications for Stability."*

40. ISO 18912-2002. "Imaging Materials — Processed Vesicular Photographic Film — Specifications for Stability."*

41. ISO 18919-1999. "Imaging Materials — Thermally Processed Silver Microfilm — Specifications for Stability."*

42. ISO 18925. "Imaging Materials — Optical Disc Media — Storage Practices."*

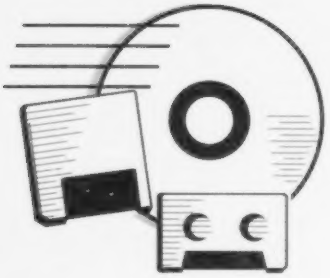
43. *Long-term Access Strategies for Federal Agencies.* Technical Information Paper, No. 12. Washington, DC: National Archives and Records Administration, July 1994.

Résumé

Un sous-comité de l'Organisation internationale de normalisation (ISO) élabore, depuis 1989, des normes sur la préservation des documents électroniques. La permanence du matériel et du logiciel constitue un enjeu d'envergure dans la préservation de l'information électronique. Il n'y a cependant pas de travail de normalisation à cet égard. Les travaux ont porté principalement sur la permanence des bandes magnétiques et des disques optiques. Des documents précisant les conditions recommandées d'archivage, en vue de prolonger la durée de vie des supports, ont été publiés pour les deux types de support. Une autre norme qui touche le consommateur a trait au traitement et à la manipulation des bandes magnétiques, et des travaux sont en cours en vue de la préparation d'un document similaire pour les disques optiques. Des méthodes d'essai ont été normalisées en vue d'évaluer l'effet de l'humidité et de la température sur les disques optiques, mais elles ne couvrent pas d'autres causes de défaillance. Un autre besoin à combler est l'absence d'essais permettant de déterminer la stabilité chimique des fonds documentaires sur support électronique.

---

* All ISO standards can be obtained from the International Organization for Standardization, Case postale 56, CH-1211 Geneva 20, Switzerland.



WAITING TO HAPPEN: LESSONS FROM PRESERVING DISASTER-AFFLICTED ELECTRONIC MEDIA IN AN ARCHIVAL COLLECTION

HANNAH FROST

Media Preservation Librarian
Stanford University Libraries
686 Escondido Road, Room A25A
Stanford CA 94305
USA
Tel.: (650) 724-4047
E-mail: hfrost@stanford.edu

Abstract

This paper describes a preservation project at Stanford University Libraries that involved attempts to recover information recorded on obsolete, disaster-afflicted computer tapes from the papers of a major figure in the history of computing. The case study explores media degradation, data recovery, disaster recovery, and digital preservation. The outcome of the project demonstrates the primary role that record creators — particularly artists, scientists, and other innovators, who, in the course of their work, develop or employ tools outside the mainstream of technology — must play in the selection for retention and preservation decision-making process. Documentation of the technical nature of media materials in archives must improve if the goal of assuring their longevity is to be achieved successfully. The experience also reveals that, as the electronic media component of archival collections increases, research libraries will need to develop staff expertise in format identification and basic media preservation tenets and techniques, including disaster response procedures.

Introduction

In December 2001, weeks after the start of my job in the Preservation Department of Stanford University Libraries (SUL), I received a message from a colleague in the Department of Special Collections. A situation had arisen involving a collection of mouldy computer tapes stored in a hot, humid mechanical room in the library basement. Could I help determine if the tapes were salvageable? This project, which evolved into a year-long foray into the new frontier of preservation — some refer to it as "digital archaeology" (Ross and Gow 1999) — is not only illustrative on matters of media degradation and disaster recovery, but also on data recovery, the history of computing, digital preservation, selection, and appraisal. The experience also lays bare the challenges research libraries are likely to face as they apply themselves to the task of preserving electronic media in archival collections.

The Engelbart computer tapes

A large store of records belonging to eminent computer scientist Douglas C. Engelbart was damaged during the 1998 El Niño season when a heavy rainstorm inundated the storage locker in which they were located. Engelbart contacted Henry Lowood, Stanford's Curator for History of Science and Technology Collections, for assistance with the disaster recovery effort; the curator jumped at the chance to acquire these records, adding to the University's existing holdings of Engelbart's papers. The affected records included paper files and approximately 225 nine-track computer tape reels from the 1980s, a period when Engelbart had been working on the Augment project in the research divisions of several corporations in Silicon Valley.

Augment grew out of the work Engelbart initiated in the 1960s at his lab, the Augmentation Research Center (ARC), located at the Stanford Research Institute until 1978. The basis of this work was Engelbart's perception that an "explosive emergence of digital technology . . . [would] cause a trend toward automating rather than augmenting peoples' activities" (Bootstrap Institute 2003). To encourage progress, it was necessary to harness and shape "the technical and non-technical elements that come into play" (Bootstrap Institute 2003) as workers operate within an organization in pursuit of its mission. Among the pioneering human–computer interaction technologies developed by the

ARC research team were the computer mouse, shared-screen teleconferencing, and hypermedia embodied in a system known as NLS (oN-Line System). One significant feature of NLS was known as the "Journal." Conceived by Engelbart to serve a similar purpose as "professional Journals and associated library services" (Engelbart 1975), the Journal provided NLS collaborators with a means to "publish" documents of any kind. Each entry was marked by an accession number upon submission; this number could be cited and linked to in subsequent submissions. In addition to fostering enhanced communication and collaboration, the Journal was intended to serve an archival purpose as a "permanent" store of records available for NLS collaborators to access at any time. Because Engelbart is credited with pioneering online computing, materials associated with his visionary work clearly have tremendous historical value.

The water-damaged paper files were shipped to a cold storage facility for freezing; the water-damaged tapes were transported to Stanford for evaluation. Upon inspection, it was discovered that the rain had not damaged all of the tapes; those spared were sent to an off-site storage facility. Among the water-damaged tapes, however, mould was discovered. To protect staff and library materials from exposure to mould, the tapes were promptly re-boxed, double-bagged, and stacked on wooden pallets in the most expedient, unobtrusive space available: a mechanical room without controlled environmental conditions in the basement of the library. In May 2000, unsuccessful attempts were made to read the tapes using campus facilities. According to e-mail records, a data recovery vendor was contacted later about the possibility of retrieving the information, but for unknown reasons the vendor's services were not retained at that time.

The task

What was recorded on the disaster-stricken tapes? The ability to answer this question and to make a responsible decision about the disposition of the materials was complicated by several underlying questions. Would the water-damaged or mouldy tapes be recoverable? Would the unharmed tapes — already 20 years old and susceptible to the degradation processes that typically afflict aging magnetic tape — be playable? What system had been used to record the tapes? Would it be possible to overcome two daunting obstacles: damaged magnetic media *and* information encoded on obsolete media using unknown systems? Did the historical value of these records justify the effort and cost of making them available?

With renewed pressure to rectify this situation, I contacted the data recovery vendor again. Together we devised a course of action: with a budget of US$500, a sample of 12 tapes (six water-damaged ones and six undamaged ones[1]) would be cleaned and analysed using the vendor's equipment. Of the six damaged items to be tested, three were mould-infested. While this was not a statistically valid sample representative of the Engelbart tape collection as a whole, it was believed that an exploratory test would help to determine if additional resources should be allocated to the cleaning and recovery of the disaster-afflicted materials.

While it is said that "most data can be rescued, if there is enough time and money" (Ross and Gow 1999), our expectations were not high. The well-known problems associated with the physical condition of tape media — binder hydrolysis caused by exposure to high humidity and lubrication loss due to evaporation or high frequency of use — would likely inhibit the success of the data recovery effort (Ross and Gow 1999). Poor tape wind was another factor, as this condition could increase the likelihood of either widespread tape damage due to particulate or chemical contamination deposited by the rainwater (Iraci 2002) or mistracking caused by tape distortion (Ross and Gow 1999). An additional consequence of the disaster, mould, presented yet another potential obstacle to data recovery (Iraci 2002). Furthermore, given the exploratory nature of Engelbart's work, there was only a remote chance that the data could be interpreted by any of the common historic computing systems the vendor had readily available for accessing obsolete formats. Nonetheless, this preliminary test was necessary to determine if the tapes had withstood the disaster and poor storage conditions to which they had been subjected and if there was any hope for making sense of the data they contained.

Discussion of test results

The test results and a description of the data recovery process and condition classification scheme used by the vendor are outlined in the Appendix. These results can be examined from two perspectives: the condition of the tape media and the ability to read the data. First, how does the condition of the damaged tapes compare to that of the undamaged tapes? Did the disaster demonstrably affect the tapes such that the ability to read the data was inhibited? And second, if any data can be read, is it possible to determine the system that recorded the data and thus interpret the contents of the files?

Of the 12 tapes tested, there was a fairly even distribution of tapes across the four classes, and there was no direct correlation between a tape's classification and water damage. While four of the tapes involved in the disaster had hard errors,[2] so too did four of the unharmed tapes. Thus, while it was expected that the disaster-afflicted materials would be more difficult to read, the results of this exploratory test did not indicate that the disaster, the mould outbreak, and subsequent exposure to poor storage conditions had a direct bearing on the ability to recover the data. However, if the disaster was apparently of little consequence, clearly media degradation presented an obstacle to data recovery. Ten of the 12 tapes exhibited varying degrees of the physical degradation typical of magnetic media of this age: the magnetic coating of the tape had become sticky, due to hydrolysis, causing difficulty in transporting the tape over the read heads of the machine conducting the analysis.

Where media degradation was not an impediment, some data were recoverable. However, the format of the data stumped the specialists. Not only were they apparently created on an unknown system, the header records[3] were like none the specialists had ever seen. A pattern emerged among the hard errors, which further piqued the specialists' curiosity. Five of the hard errors occurred at the same point: file 1, block 202. Did this indicate the data were compressed or encrypted during the backup process? Or was there some other characteristic of the original computing environment that could explain this pattern?

Further probing and resolution

All parties agreed that the next step was to find answers to questions concerning the technical nature of the tapes and the context of their creation. A series of detailed questions compiled by the data recovery technicians was forwarded to Engelbart. He in turn contacted a colleague, Raylene Pak, who had been involved in the research projects that the tapes documented and now had access to existing records about them.

Pak provided important information about the tape collection, including the system hardware, the operating system and patches added by project staff, how and when the various system components and programs were backed up, and how to interpret the tapes' label notations in order to identify their contents. Despite this information, however, there was no apparent explanation for the consistent hard error at file 1, block 202, except for possibly one clue. Pak writes, "In going back through my old email and documents pertaining to tapes, I find several references to difficulties in restoring files from tape. Evidently, this type of problem existed even when the tapes were relatively new" (Pak 2002). The colleague also relayed that "[Engelbart] is most interested in getting copies of old Journal document files back from tape. The primary source for such files would be the Journal archive tapes. However, full and incremental dump files may also be a source" (Pak 2002).

With this new information, the curator determined that any Journal archive tapes, which document what he believes to be the earliest known virtual community, should be retained. Engelbart's confirmation of their historic value and Stanford's existing holdings of paper copies of Journal files spanning a period of 3 years, materials frequently used by researchers today, justified this decision.[4] These files would be in the Augment file format, yet with a skilful programmer and available documentation, it was possible that the Journal files could be read without the use of the Augment system itself. While SUL is in no position to re-create the Augment system, and has no plans to do so in the future, the curator was aware of other institutions that had demonstrations of the system running. Furthermore, Engelbart's current work[5] could make use of the early Journal files, if the opportunity, will, and resources to recover them in earnest should become available in the future.

I searched through the boxes, carefully deciphering the label notations in search of Journal archive tapes, and found none. The water-damaged tapes were discarded; the remaining undamaged Engelbart tapes and those from other accessions will be reviewed and weeded in the coming months when the collection is prepared for relocation to Stanford's new high-density collections storage facility.

The end?

As the preservation field develops a body of knowledge around the longevity and deterioration mechanisms of electronic media, it is tempting to use a case study such as this one to draw a tidy conclusion about digital media endurance with respect to disasters or system obsolescence. Of the possible factors — typical magnetic tape deterioration, media deterioration caused by water damage or mould, errors when the data were originally written, or the lack of appropriate computing systems — would it not be useful to know if one factor trumped the others in preventing the recovery of the data? Yet the variables were too numerous, and the costs to eliminate them too high given limited resources, to draw a sound conclusion based on empirical testing. While it was true that the aging tape media, disaster-afflicted or otherwise, had reached critical levels of deterioration, the only claim that can be made with confidence is this: ultimately it was the inability to re-create the original computing environment that became the primary obstacle to accessing the materials.

While the testing process did not result in recovered data, we were able to answer many of the questions raised by the Engelbart tapes. We now know how to identify the tapes with content of historic value, that the damaged tapes contained no Journal archive files, and that media degradation and errors in writing the datatapes present potential obstacles to data recovery, but that data recovery is possible. Yet it remains unknown whether or not the Stanford collection contains any Journal archive files, and in this respect the outcome of the project remains unresolved.

Unexpected discovery

The story does not end here, however. While searching for Journal archive tapes among the scores of damaged media, I uncovered a surprise: seven reels of motion picture film and two videotapes. These materials had clearly been exposed to silty water in the 1998 disaster. The box listing had not indicated that any materials other than computer tapes were part of this collection. Apparently, in the rush to confine the mouldy materials temporarily until some time in the future when the situation could be properly addressed, no one had noted the existence of the analog audiovisual media.

SONY
PROPER CARE OF VIDEO TAPES
INDEX

This small bag containing information on video care was found in a case alongside a ¼-inch reel of video recorded by the ARC project team. The rust and silt residue dried onto the bag clearly indicates that rainwater contaminated the video during the disaster.

Two film labs evaluated the condition of the motion picture footage and reported that it was not infested with mould, though some film items had suffered damage as the unclean water dried onto the emulsion surface. The films have been cleaned, cored, and re-canned, and the videotapes cleaned, re-housed, and reformatted. The films and videos contained visual documentation of some of the ARC lab's inventions such as the NLS in operation, and had been presented by Engelbart at several conferences throughout the 1960s and 1970s. According to the curator, these items are "outstanding companions to Engelbart's famous 1968 demo [of the computer mouse, of hypermedia, and videoconferencing], which is one of the key pieces of documentation on the early development of human–computer interaction anywhere" (Lowood 2003). If it is never possible to access any Engelbart computer files retained in Stanford's collection, these documentary materials may be loosely considered as a form of migration (Wheatley 2001), one which may prove to be the only effective and feasible means to preserve a sense for the original "look and feel" of Engelbart's highly influential contributions to the history of computing.

Lessons learned

It is the role of cultural heritage institutions to preserve the works of society's great minds. Artists, scientists, and other innovators, who, in the course of their work, develop or employ tools outside the mainstream of technology, present the caretakers of their records with tremendous challenges. It is precisely the unusual, pioneering nature of work like Engelbart's NLS Journal that heightens scholarly interest in it, and, at the same time, increases the difficulty in maintaining its viability, authenticity, and integrity. As libraries continue to build research collections, the electronic media component of the collections will inevitably and steadily increase. It is therefore useful to consider the lessons learned from the Engelbart case study in order to inform and influence the way institutions approach the task of preserving archival media materials and to set reasonable expectations about their success.

The first lesson concerns media awareness. Cultural heritage institutions that collect contemporary materials must develop staff awareness about electronic media formats and techniques used in their preservation. The Engelbart tapes may have been destined for trouble from the moment the data were laid down, but clearly there are shortcomings in the way this situation involving water-damaged media materials was managed.[6] Staff handling the damaged materials knew what to do for the water-logged papers, but as for the materials in cans, they threw up their hands. Ideally the media materials should have been promptly rinsed in distilled water and dried at room temperature; if these measures had been taken, the analog items would have been identified and separated from the digital tapes. That the analog audiovisual materials among the datatapes were not identified when the materials were removed from the water-damaged storage boxes and re-boxed for temporary storage indicates a lack of awareness about the distinctions between electronic media types and their preservation needs. Barring extreme damage or deterioration, to clean, playback, and copy decades-old video and film for preservation purposes is a significantly simpler process than doing the same for nine-track datatapes from the same era. The ability to distinguish the formats and features of media materials is not only critical in disaster response, it is also fundamental in accomplishing routine archival management activities, such as administrative record keeping about collection acquisitions, processing for public access, and preservation planning.

The second lesson concerns the role of information gathering in the appraisal of digital materials for archival retention and in the selection of digital materials for preservation attention. In the process of working with the data recovery vendor, it was the crucial background information provided by the individuals involved in creating the records that brought the project to resolution. It was not necessary to access and read the data itself to determine whether or not they were worth saving. Instead it was the technical and contextual information provided by the records' creators that enabled an informed selection decision-making process to occur. If this key information about the data recordings had been sought and obtained at the time of their rescue, then the items deemed of historic value could have been culled, simplifying the disaster response effort and preventing the subsequent expenditure of scarce preservation resources as well as the potential dangers presented by the mould.

The essential role of documenting the technical characteristics of digital materials is a recurring theme within the literature on digital preservation. As Jones and Beagrie (2001) state, "there are factors which make documentation particularly critical for the continued viability of digital materials and they relate to fundamental differences between traditional and digital resources." Preservation metadata embodying the technical and other contextual information about a digital object are considered key to the successful long-term management of data within digital repositories like those under development in the library community. In related work underway within the contemporary arts conservation community, such as the Guggenheim Museum's Variable Media Initiative, the recording of technical and contextual details about electronic art, new media, installation art, and other ephemeral manifestations of cultural production plays a similarly key role. Yet it is less clear to what extent this practice is being widely adopted for the care of digital materials stored among analog records within manuscript collections, organizational records, and other bodies of archival documents. Careful documentation of the technical context for these works must occur, and must occur early in the acquisition process, if their preservation is to be successful.

SUL has begun to integrate the principles of media awareness and technical documentation into the management of the electronic media materials in its archival collections. The Media Preservation Department is collaborating with Special Collections staff on an assessment of the collections. Using a survey approach, individual electronic media materials are evaluated. Analog items (videotapes, sound recordings, films) are assessed for their physical condition, age, and any vulnerability associated with the particular format, media, or the nature of the recording itself. The gathering of this information enables collection caretakers to achieve an unprecedented, highly detailed understanding of the nature, condition, and extent of these materials within Stanford's existing and incoming archival holdings. The availability of this valuable information facilitates the process of identifying preservation reformatting priorities and builds awareness among staff across the organization of the needs of archival media.

Digital materials are assessed similarly in preparation for their preservation in the Stanford Digital Repository. A core team of librarians is developing a process that aims to identify and document the features of digital objects, including those found in archival collections, to be preserved in the repository. Features that are deemed integral to the object's usefulness and value as an information resource, from a curatorial perspective, are recorded. Technical features of the object likely to be susceptible to change or loss as the object is migrated are evaluated in terms of their relative complexity.[7]

The interplay of these characteristics is considered in order to determine the level and extent of preservation service provided by the repository over time. It is believed that these new strategies underway at SUL will guide the successful care of its electronic media materials well into the future.

Acknowledgements

The author is deeply grateful to her collaborators and supporters on this project: Michael Olson, Connie Brooks, Walter Henry, Catherine Aster, Henry Lowood, Steven Mandeville-Gamble, Ron Nakao, Lorren Livingston, Chris Muller, and Douglas Engelbart.

Endnotes

1. According to the label information, each of the 12 nine-track tape reels had been used repeatedly over the course of several years throughout the 1980s. Tape 1 is an unknown brand. Tapes 2–11 are various formulations of Memorex datatape stock, including Super Reel, MRX III, IV, and V; Tape 12 is BASF Endura stock.

2. A hard error is one associated with hardware, such as a disk or tape read error, which cannot be recovered or processed.

3. A header is the first record in a disk or tape file. It typically contains basic file identification or structural information.

4. These paper copies of Journal files, housed in binders, had been created and were actively used by Augment project staff. Although it is not altogether clear why they were produced, it is possible that resource limitations or technical difficulties hindered the ability to write all files to tape.

5. For information on Englebart's current work, as well as a more complete picture of his past research and accomplishments, see the Bootstrap Institute's Web site (www.bootstrap.org/).

6. Circumstances surrounding the acquisition of the Engelbart tapes bear mentioning for they provide the necessary context from which subsequent events played out. In February 1998, SUL experienced extensive flood damage in four of its buildings due to severe rains caused by the weather phenomenon known as El Niño. Approximately 80 000 books were affected by the flooding. It was not until late summer 1998 that these materials were completely treated and returned to the library stacks. During this unsettled period, the damaged Engelbart materials arrived at SUL; at the same time another collection arrived — a voluminous and prominent collection associated with the founder of the University — which had been donated to Stanford with only 1 week to arrange for its transport and accession, a process that consumed the attention of collection management staff. In the flurry of activity, it appears that subsequent miscommunication, coupled with a lack of understanding about the materials' potential research value, resulted in the neglect of the mouldy Engelbart media materials. Further complicating this situation, the institution had never before been faced with the daunting task of preserving more than 200 damaged, obsolete digital media materials. Indeed the notion of digital preservation as a service and function of research libraries was still in its infancy in 1998. Given this context, it is no wonder the matter of the mouldy media went unresolved for 4 years.

7. The approach developed by SUL is influenced by the following work: Bennett, J.C. *JISC/NPO Studies on the Preservation of Electronic Materials: A Framework of Data Types and Formats, and Issues Affecting the Long Term Preservation of Digital Material*. British Library Research and Innovation Report, No. 50. 1997. www.ukoln.ac.uk/services/elib/papers/supporting/ pdf/rept011.pdf (accessed 5 May 2003)

Bibliography

Engelbart, C. "A Lifetime Pursuit: From a Biographical Sketch of Douglas C. Engelbart." Bootstrap Institute, 2003. www.bootstrap.org/chronicle/chronicle.html (accessed 16 October 2003)

Engelbart, D.C. "NLS Teleconferencing Features: The Journal, and Shared-screen Telephoning." www.bootstrap.org/augdocs/augment-33076.htm (accessed 16 October 2003). Originally published in *COMPCON '75 Digest: Proceedings of the COMPCON Conference, 9–11 September 1975*, 4 pp.

Iraci, J. *Disaster Recovery of Modern Information Carriers: Compact Discs, Magnetic Tapes, and Magnetic Disks*. Technical Bulletin, No. 25. Ottawa, ON: Canadian Conservation Institute, 2002, 15 pp.

Jones, M., and N. Beagrie. *Preservation Management of Digital Materials: A Handbook*. London, UK: British Library, 2001, 139 pp.

Lowood, H. E-mail correspondence with the author, 4 February 2003.

Pak, R. E-mail correspondence with the author, 15 August 2002.

Ross, S., and A. Gow. *Digital Archaeology: Rescuing Neglected and Damaged Data Resources.* London & Bristol, UK: British Library and Joint Information Systems Committee, 1999, 81 pp. www.ukoln.ac.uk/services/elib/papers/supporting/pdf/p2con.pdf (accessed 11 February 2003)

Wheatley, P. "Migration – A CAMiLEON Discussion Paper." *Ariadne* 29 (2 October 2001). www.ariadne.ac.uk/issue29/camileon/intro.html (accessed 11 February 2003)

Appendix: Data recovery process, condition classification scheme, and test results

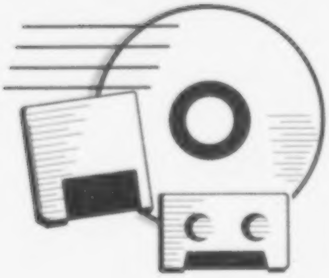
The data recovery process begins by loading a tape on a machine to attempt to read recorded data on the very first pass. Software designed by the data recovery vendor analyses the data in search of clues (e.g. the presence of file headers, packed fields, and whether it is ASCII or EBCDIC text code) to indicate how the tapes were originally written. The read attempt is aborted if there is evidence that the tape is dirty or gummy; if this is the case, the tape is loaded on equipment that cleans and re-tensions the tape as well as identifies and verifies "bad spots." Once clean, the read process is resumed. In the course of reading, the condition of the tape is classified using a scale developed by the vendor. A Class 1 tape is one that runs cleanly through the read process. Class 2 defines those tapes that may exhibit a soft or hard error, but the read attempt is largely unaffected. A Class 3 tape has hard errors in multiple areas where tape damage, due to either a poor recording originally or subsequent media degradation, prevents data access. The worst state, Class 4, describes those tapes where the magnetic coating becomes detached as the tape is read, causing "voids" or blank spots on the tape. In some cases the voids are small, sporadic, and can be easily skipped over; in other cases, the entire tape is unreadable. "Shoe shining" is a term used to describe the behaviour of a tape drive that encounters a Class 3 or 4 tape. When the drive encounters a damaged area, it will make repeated attempts to read it. In each attempt, the tape is advanced, stopped, reversed, and read. This process may be repeated up to 30 times. This back-and-forth action resembles the movement of a person shining a shoe.

Table 1. Tape cleaning and data recovery test results
(courtesy of Muller Media Conversions, Inc.)

Tapes			Class					1st hard error	
			1	2	3	3–4	5	File	Block
Disaster afflicted	Mouldy	1		X				1	202
		2					X	1	202
		3		X				1	202
	Not mouldy	4	X						
		5		X					
		6					X	7	3843
Unharmed		7	X						
		8				X		0	
		9			X			194	9736
		10			X			1	202
		11			X			1	202
		12	X						

Résumé

Nous décrirons un projet de préservation en cours dans les bibliothèques de l'Université de Stanford dans le cadre duquel on s'efforce de récupérer des bandes informatiques périmées et abîmées faisant partie de dossiers d'un personnage majeur dans l'histoire de la micro-informatique. Cette étude de cas est l'occasion de s'interroger sur la détérioration des supports, la récupération des documents abîmés et la préservation numérique. Le résultat du projet confirme le rôle fondamental que jouent les créateurs de documents (surtout les artistes, les scientifiques et autres innovateurs qui, dans leur travail, élaborent ou utilisent des outils spécialisés) dans le choix des documents et le processus décisionnel. Il faut améliorer la qualité de la documentation relative au caractère technique des supports dans les archives si l'on veut garantir leur longévité. L'expérience révèle également que, à mesure que le nombre des documents électroniques augmente dans les collections, les bibliothèques de recherche devront perfectionner leurs employés en matière de techniques d'identification de format, de préservation des supports et, notamment, d'intervention en cas d'urgence.



LONGEVITY OF CD MEDIA: RESEARCH AT THE LIBRARY OF CONGRESS

CHANDRU J. SHAHANI, BASIL MANNS, AND MICHELE YOUKET

Preservation Research and Testing Division
Library of Congress
Washington DC
USA
E-mail: csha@loc.gov

Abstract

This paper discusses some investigations at the Library of Congress into the natural as well as accelerating aging of audio CDs. The natural aging study was designed to monitor the long-term effects of real-life storage and use conditions on the integrity of a limited sample of audio CDs selected randomly from the Library's collection. During the course of this ongoing investigation, the sample population has been retained within the collection and has remained subject to normal wear and tear; it has been withdrawn only briefly every few years for testing. Results from the past 7 years of this pilot project are presented, along with implications for its future course. Also presented is the status of an accelerated aging study of audio CDs that is currently in progress, as well as visual observations of defects generated by the accelerated aging process.

The goal of these studies is to gain an understanding of the scope of the challenge that may lie ahead for the long-term preservation of a contemporary CD collection that must inevitably comprise a wide range of compositional characteristics in every layer of its manufacture. It is hoped that this research will ultimately enable us to 1) provide the best long-term protection for our collections, 2) establish relevant parameters for monitoring CDs that would signal the need for creating backup copies or migrating the data, and 3) identify the failure mechanisms that arise from deficiencies in the manufacture of CDs.

Introduction

Optical storage media are not subject to the same wear and tear as other recorded media, and it has sometimes been presumed they have the potential to last indefinitely. Hence they have had a strong appeal for preservation-conscious institutions. For this reason the Library of Congress has long been committed to gaining an understanding of the lasting qualities of optical storage media.

The earliest studies at the Library date back to the mid 1980s. At that time an accelerated aging project was undertaken with 12-inch optical WORM discs to investigate the application of the Arrhenius aging methodology for the determination of their end-of-life point. Specific test procedures were developed for these analyses and formalized in American National Standards Institute (ANSI) standard X3.199-1991.[1] Later, with support from the National Institute of Standards and Technology (NIST), the Association for Information and Image Management (AIIM) developed an error-monitoring scheme for WORM, magneto-optical (MO), and CD optical media.[2,3] These tests involved accelerated aging of the discs in an elevated temperature and humidity environment, typically for repeated periods of 500–2000 hours, to accelerate potential chemical changes that may affect the operational characteristics of the discs. At about the same time, IT9 (an ANSI standards committee) was developing life-expectancy procedures for digital optical media and approved three standards dealing with CD-ROM media.[4]

The primary objective of the current research is to assess the preservation needs of the audio CD collections at the Library of Congress, and devise strategies to minimize any data loss that may result as the collections age. The first step in this process is to gain an understanding of failure modes by monitoring errors that accumulate on aging CDs. New insights might also be gained through a scientific examination of any visually perceptible defects. It is our hope that we will ultimately be able to link compositional characteristics of individual CDs to failure mechanisms in a way that will enable us to flag vulnerable CDs for replication or backup before any data are lost.

The current project involves studying the effects of the aging process on CDs under natural as well as accelerated aging conditions. The accelerated aging study was modeled after the ANSI standard mentioned above.[4] Although these experiments are still in progress and some of the data sets have yet to be completed, several interesting observations have been recorded about the development of some clearly visible physical defects. The nature of these defects and their impact on the playability of the discs is presented in this paper. However, as with most accelerated aging experiments, this series of experiments also begs the question of whether or not the failure modes observed are truly representative of phenomena that occur in real life, and if the differences that emerge in the aging of CDs with different manufacturing characteristics will ultimately stand the test of time.

In anticipation of these inevitable doubts, a study to monitor the natural aging of a representative sample of CDs was initiated in 1996. Given the limited resources available to us at that time for testing CDs, this initiative was restricted to a pilot study of 125 discs selected randomly from the Library's audio CD collection, which in 1996 was estimated to consist of 60 000 items. While this was not a statistically significant sample, its size was deemed adequate to justify a full-fledged natural aging study if significant data errors were to develop in even a few discs. Indeed, that is exactly what transpired (as can be seen from the data presented below). Since the study was initiated, the size of the Library's audio CD collection has mushroomed from 60 000 to 380 000. In light of this growth, a statistically valid random sample is now being assembled for a more definitive natural aging study.

Natural aging study

Experimental protocol

Test population: At the beginning of the study, a representative test population of 125 CDs was randomly selected from the Library's collection. These CDs were periodically withdrawn for a brief time for the purpose of testing; otherwise they were left to experience the same wear and tear as the rest of the collection. During the course of the study the test population was reduced by 6 CDs (these could not be retrieved, most likely because of removal of duplicates from the collection).

Testing methodology: A specialized tester, the CD-CATS/SA3[5] (designed specifically for CDs according to the Philips Red/Blue/Green/Orange Book standards, ISO9660, and extensions[6]), was used to read and test the CDs for 25 test parameters according to the physical specifications of CDs as defined in ISO 10149.[7] In addition to the ISO specifications, the tester also measured jitter and effect of length deviation based upon production level requirements of Philips and Sony. Definitions of these test parameters along with the ISO documents are accessible in published literature,[5,7] and their acronyms and brief definitions have been included in the Appendix at the end of this paper. It is generally accepted that these measurements and statistics are good indicators of CD operability.

The test results are grouped by the CD-CATS system into three data categories: static data, dynamic data, and the recommended length deviation and jitter data. BLER, E32, and jitter are perhaps the three most widely used indicators of errors or potential problems, although many other parameters also give an indication of undesirable changes.

The tester also generates an "OUT-OF-SPEC. LIST" based on the maximum or minimum values for each parameter as specified by the user of the CD-CATS system, a useful reporting mechanism for flagging or monitoring levels or errors. However, an out-of-specification level for a parameter does not necessarily indicate an error, as it depends on the nature of the parameter and where it occurs on the disc. A simple example of the irrelevance of an observed "error" is one that occurs on an area of the disc where no data are recorded.

Testing frequency: Testing was first conducted in 1996 (Run 1) to establish the baseline test parameters for each of the 125 CDs. This testing of individual CDs was repeated in 1999 (Run 2) and again in 2003 (Run 3). We plan to continue this testing every 3 years.

Results and Discussion

When digital optical media first became available, it was commonly believed they would be free of the wear and tear that traditional media suffered every time they were played. However, now and then there were cases where CDs were rendered unplayable due either to some unintentional abuse (e.g. the penetration of an adhesive from a label, or penetration of the protective coating under a label due to writing on it with a pen) or to a defect in their construction that somehow made them more vulnerable to the aging process. Therefore, in 1996, when the Library's collection numbered about 60 000 audio CDs, we decided to initiate a program to monitor the condition of a representative sample of CDs to explore the possibility of error generation in them. The selected CDs would remain in their place in the collection and be subject to the same probability of being played as the rest of the collection. Every few years they would be withdrawn from the collection to be tested in a non-destructive manner.

Back-of-the-envelope calculations in 1996 suggested that a random sample of 125 CDs would be adequate to gauge the aging process of the collection within a 90% confidence limit. The limited resources available to us at the time for testing individual CDs was also a factor in limiting the size of the test population. However, within a few years it became evident that the Library's CD collection was growing exponentially and the size of the test sample would need to be increased appreciably for it to be truly representative. But having already embarked on the project with a test sample of 125, we decided to simply limit our expectation of the project rather than abandon it and start over. Hence we reduced the study to the level of a pilot project, the results of which would be used to decide if a full-fledged study with a statistically valid sample would be justified.

Figure 1 represents the 125 sample discs in terms of their copyright dates or production dates as specified on the CDs themselves, their covers, or in the literature accompanying them. These dates may not be the exact dates of manufacture, but they are the best dates that can be associated with the given CDs. For all practical purposes we can assume that this sample is a good representation of the age distribution of the Library's CD collection in 1996. The increasing sample density in the later years, which is clearly evident from the representation shown here, is expected given the limited use of CD media at first, their gradual acceptance, and their ultimately dominant position in the music publishing world.

Examination of the test data: All test parameters were collected from the CD-CATS system for each of the 125 CDs. All CDs were run at X1 speed and 100% sampling was used. To date, three successive tests have been conducted, Run 1 (baseline, 1996), Run 2 (1999), and Run 3 (2003). The absolute values of these parameters naturally include the impact of the manufacturing process as well as the effects of aging and physical handling prior to the time our study was initiated. Because our concern is limited to the effects of natural aging from that point onward, the importance of the values of the measured parameters in this study is their cumulative change from the baseline value rather than their absolute value at anytime.

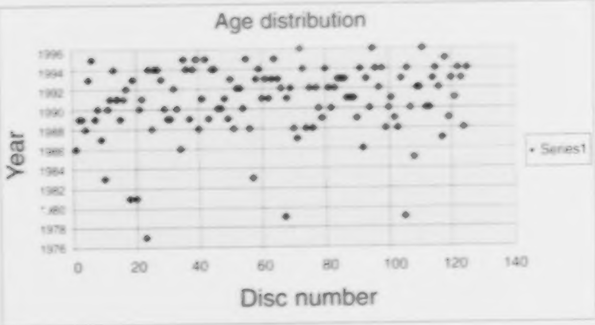


Figure 1. Approximate age distribution of sample.

All error functions were tabulated and used to draw inferences in comparison to other errors, such as the correlation between BLER and E32. Also of interest were the parameters that changed with time. In all cases, before any judgment could be made as to an "error," the specific measurement had to be carefully examined as to where it occurred and what other parameters indicated an error at the same time. This is necessary because discs may be manufactured with certain abnormalities on a particular part of a particular track that will show up as errors, but will have no effect on the playability or reading of the useful data.

BLER is often considered a good general indicator of the condition of a disc. Many times there is a correlation between BLER and some of the other test parameters mentioned above, although sometimes there are unexplainable exceptions to this fact. A scatter plot of the BLER for Run 1 (the baseline run) of the sample discs is shown in Figure 2. Scatter plots of the BLER for Run 2 and Run 3 follow in Figures 3 and 4, respectively.

It is common knowledge that the technology for manufacturing CDs underwent extensive improvement over the years that encompass the period of this study (which includes all CDs manufactured up to 1996). Given this fact, it is interesting to note that some of the earliest CDs are lasting very well. For example, for disc 23 (which has an associative date of 1977), the BLER remained relatively constant during the three test runs, being 36, 34, and 36. Likewise, discs 66 and 105 (both of which also date back to the late 1970s) exhibited little change in the BLER. Thus, some of the CDs produced with the earliest technology are holding up very well 25 years or more after their manufacture.

While BLER has been highlighted in this discussion, and the only scatter charts presented are for this parameter, it is obviously not the sole parameter that

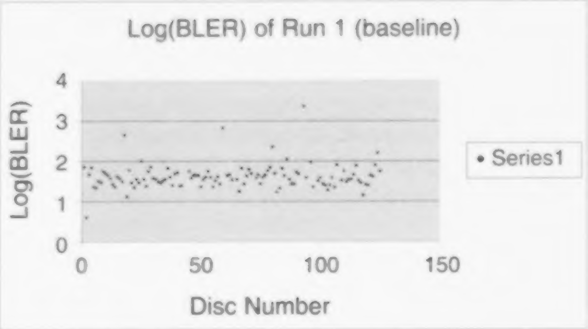


Figure 2. Plot of log(BLER) for Run 1 (baseline, 1996).

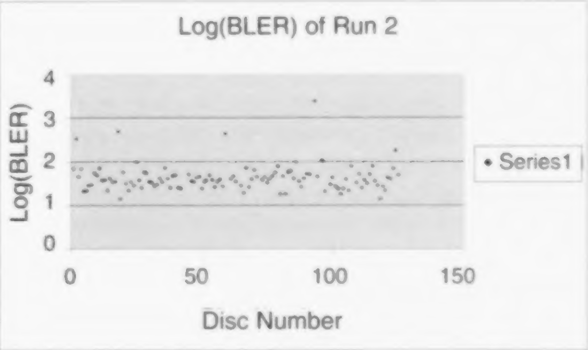


Figure 3. Plot of log(BLER) for Run 2 (1999).

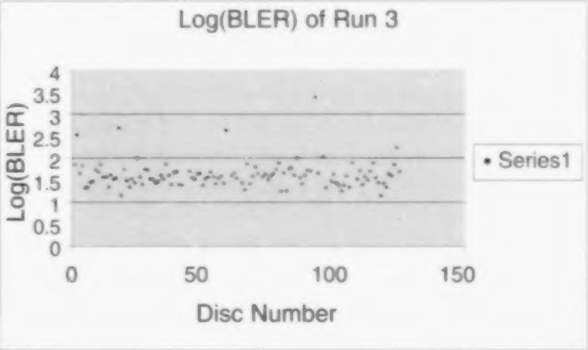


Figure 4. Plot of log(BLER) for Run 3 (2003).

merits attention. Scatter plots were also produced for all the other parameters in order to discern if there were any correlations between parameters. Relationships between parameters can be instructive for an individual disc where, for example, the BLER may be high but the E32 low, or where the BLER and E32 occurred at about the same time. Surface presentation can be used to determine the location of errors on the disc. These types of analyses were conducted but, because of the sheer volume of data generated, they are not included here.

Four discs (numbers 18, 59, 80, and 93) had a BLER greater than the generally acceptable maximum rate of 220, or log(BLER) of about 2.34. These same discs also exhibited E32 errors, i.e. fatal uncorrectable errors. However, there was no noticeable audio defect in the recordings — all four of the discs were playable. If any of the errors were in the data sectors, they were evidently correctable.

Figure 3 shows the results after Run 2. Note that the BLER for disc 2 went from 4 to 355, while the BLER for disc 80 went from 228 to 18. The latter change is hard to comprehend. The only rational explanation that occurs to us is that a spec of dust residing on the reflective side of disc 80 might have caused the initial high BLER reading, and the subsequent removal of this spec after routine cleaning (each of the discs was cleaned using an isopropyl alcohol wipe before introducing it in the CD-CATS tester) was responsible for the decline. Other parameters for disc 80 remained about the same or normal for a typical measurement. Discs 18, 59, and 93 showed little change and were all in excess of the standard limit for all three runs. Discs 18 and 59 had readings of BLER in the mid 400s while disc 93 had a BLER of about 2300.

The BLER data by themselves to not reveal any significant variations from successive runs, except for discs 2 and 80. As already indicated, we have ascribed the downward change in BLER for disc 80 to an experimental aberration. However, the rise in BLER for disc 2 is real. Figure 5 shows the average BLER for the entire sample for each of the three runs. The upward drift in the average value of BLER does indicate a worsening trend. Further measurements scheduled for 2006 will be valuable in lending definitive confirmation or denial of this trend.

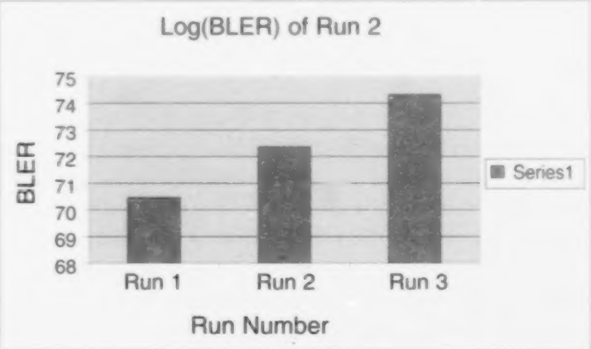


Figure 5. Average BLER for three runs.

Another parameter analysed was normalized reflectivity. The scatter plot for this property was fairly flat and only two CDs had out-of-specification values. However, these two CDs did not have any E32 errors and their BLER values were within specification, suggesting that the reflectivity "errors" were not significant but were worth observing for later runs. There are different ways of calculating this parameter and it is often an indication of surface dirt or scratches, which will not always result in a fatal error.

Radial noise was also examined and all CDs were within specification except disc 68, which had no other out-of-specification parameters. A high value of radial noise can occur due to poorly defined pits during production, surface dirt or scratches, or other factors (such as birefringence) affecting or interfering with the return signal. Again, the importance of this parameter is only in the change in its value and what role it may play in causing errors in future years.

Finally, maximum length deviation distributions and maximum jitter distributions on both the land and the pit region were monitored. These are critical test parameters to indicate the error levels on CDs. In this study, whenever an E32 error occurred along with high BLER levels, the land length deviation and jitter maximums were also at out-of-specification levels. This gives good correlation with expected results.

Table 1 tabulates all errors, or parameters above pre-established levels, for all three runs. The numbers represent the number of times the test parameter exceeded the pre-defined level, as a percentage of the total sample. The pre-defined levels are specified in ISO 9660[6] and ISO/IEC 10149.[7] The parameters are explained further in the Appendix at the end of this paper.

It is important to note the significant increase in length deviation and jitter in the last run. Length deviation is the length of each pit and land area measurement compared to the ideal for each time interval, 3T to 11T. Jitter is measured individually for each pit and land 3T to 11T transition. Jitter is measured for the entire track, then calculated and expressed statistically, as a standard deviation.

Bit error can be predicted based on length deviation, signal jitter, and player jitter. Low jitter does not always indicate high quality, but can indicate low average error rates. Jitter does not indicate local defects, such as black spots, birefringence, or scratches.

Table 1. Percentage of CDs at out-of-specification levels for the three runs

Type of measurement	Percentage of CDs out-of-specification		
	Run 1 (baseline)	Run 2	Run 3
Static data			
SLD		2.5	
SPD		3.3	2.5
MID		0.8	
SVY	3.3	4.2	4.2
TRP	1.6	4.2	
ECC		0.8	
Vertical data			
DEV			
DEFL			
PP	8.8	15.8	8.3
PPC	17.6	40	20
XT		0.8	
SVY	12.8	9.2	6.7
Dynamic test data			
BLER	2.4	3.3	4.2
E11	1.6		
E21	0.8		
E31	0.8		0.8
E12	0.8		
E22	11.2		
E32	4.8	4.8	4.8
I3			
I11	2.5	4.2	2.5
I3R			0.8
I11R	1.6	2.5	1.6
REF		3.3	
SYM	1.6	9.2	7.5
BERL	8.8	10.8	13.3
CRC			
RN	4.8	10.8	7.5
Time interval data			
Length deviation		7.5	100*
Jitter	57.6	15	100*

There are other parameters and specific changes that can be evaluated on a case-by-case basis. Physical examination using a microscope is also important to look for surface defects or other physical abnormalities. Such analyses are in progress, but no definitive observations are available for presentation at this time.

Accelerated aging studies

Experimental protocol

Accelerated aging was based on ANSI NPM IT9.21-1996.[4] The test conditions are summarized in Table 2. Eighty CD-ROMs were used in each batch, and each batch was divided into five groups. Each group was tested for BLER and all the other parameters, then placed in the stress chamber for the specified time, and then re-tested. This process was repeated 4 times until the total soaking time was achieved. The sampling size was sufficient to arrive at a cumulative failure distribution, using the Weibull Distribution.[4] The Eyring Model[4] could then be applied to generate a life expectancy for the sample.

Table 2. Test conditions for accelerated aging

Test condition	Stress test time (hours)	Total soaking time (hours)	Number of CDs
80°C @ 85% RH	500	2000	10
80°C @ 70% RH	500	2000	10
80°C @ 55% RH	500	2000	15
70°C @ 85% RH	750	3000	15
60°C @ 85% RH	1000	4000	31

Results and discussion

The accelerated aging experiments are still in progress. However, even after the first 500 hours of accelerated aging under the conditions specified above, some significant failure modes have been observed. These failures, which seem to become more significant with increasing aging time, can be roughly classified into three major types.

The first type of failure is spots appearing on the CD. There are a number of different types of spots — black, white, and bubbles — not all of which we understand at present. Some seem to form on the polycarbonate or reflective surface, some within the surface, and some on the metallized surface, either on the reflective side or causing aberrations on the label side. Figure 6 shows a disc with a conglomeration of such spots. These spots were not formed by the condensation of water within the CD structure (as can happen if the relative humidity within the aging chamber is not slowly ramped up when aging is initiated and ramped down just as gradually before the termination of an aging period), and are therefore the result of the aging process itself. A magnified view of a spot on the metallized layer is presented in Figure 7, while Figure 8 shows how such chemical degradation can wipe out recorded data.

The second type of failure is the visual disappearance of the metallized layer (see Figure 9). In this instance the reflectivity of the metallic aluminum or aluminum alloy layer of the disc is lost both to the eye and to the read laser. The only logical explanation that suggests itself for this transition is that the edge of the disc was

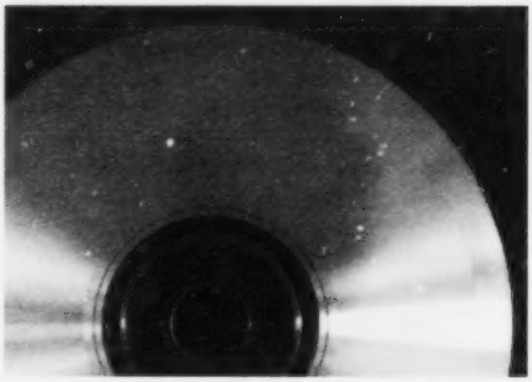


Figure 6. Disc showing spots after accelerated aging.

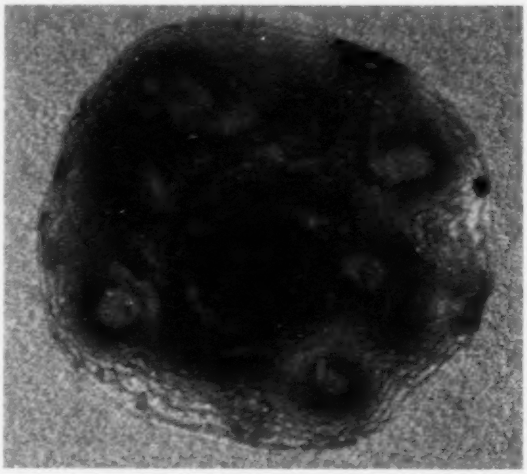


Figure 7. Closeup view of a spot on the metallized layer of a disc after accelerated aging.

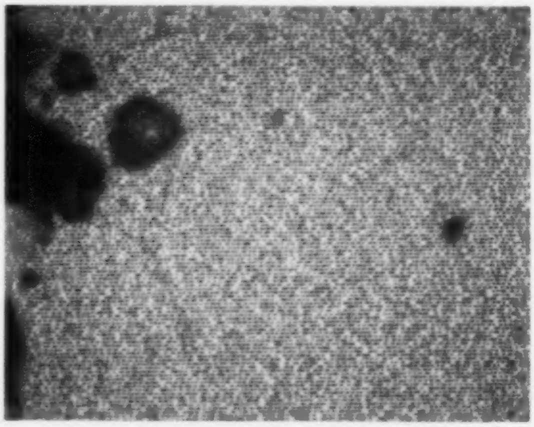


Figure 8. Closeup view showing the effect of a spot on recorded data.

not sealed properly during the manufacturing process or the seal was somehow compromised during the aging process, thus rendering the aluminum vulnerable to oxidation. This phenomenon occurs gradually, and only time will tell if it will also happen under natural aging conditions — although the possibility appears real. It was noticed on only one type of disc. Interestingly enough, some of the printing on these discs appears to have exerted a protective effect since the reflectivity under the printed image was retained. It is hard to imagine how the printing provides a seal against the environment in these areas, but we are not in a position to offer a more logical cause for this observation at this time. This visual disappearance of the metallized layer, even if it is real and also happens under natural aging conditions, will be limited to CDs of early manufacture as current manufacturing technology provides a much better seal all around the CD.

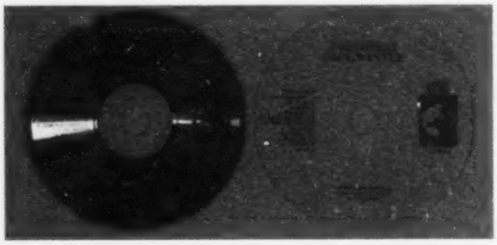


Figure 9. CDs before (left) and after (right) accelerated aging at 80°C and 70% RH for 500 hours. Note that the metallized layer seems to have disappeared on the aged CD.

The third significant type of failure is the peeling or flaking off of the metallized layer when a CD is cycled between hot and cold temperatures. Figure 10 provides an example of a CD experiencing both peeling and flaking off. This example is not an isolated case, although it is a rather extreme one of its kind. This particular CD was exposed to both hot and cold temperatures. Although the IT9 standard does not address cold temperature cycles, this experiment was undertaken to satisfy our curiosity about the effect of the extreme temperatures a CD might experience under circumstances such as being played in a car in winter time. In some rare cases, we also observed similar, although more limited, damage when discs were subjected only to elevated temperatures, i.e. no temperature cycling. It seems that the resin protective layer shrinks and, because it is strongly attached to the metallized layer, causes the metallized layer to peel off the polycarbonate substrate.

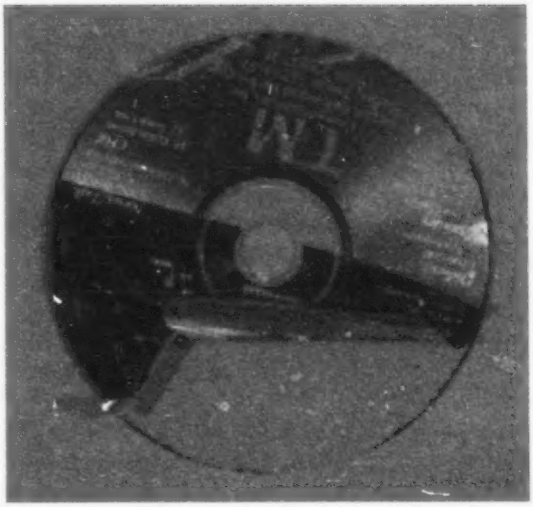


Figure 10. The metallized layer of this CD is peeling and flaking off the polycarbonate substrate.

Conclusion

Although the natural aging pilot project has been in progress for only 7 years, the limited data obtained thus far have demonstrated a slight increase in the BLER and a significant increase in the length deviation and jitter measurements. But these measurements cover only the CDs that were in the Library's collection in 1996, and the size of this collection has grown exponentially since that time. Therefore, a new natural aging study is being initiated — a study that is justified by the rising trend in BLER observed in the current study. In the new study, the sample size will be increased to more than 1000 CDs so as to obtain data with a 95% confidence limit in its representation of the entire collection of 380 000 discs.

The accelerated aging studies of CD-ROMs (mostly audio) have been time-consuming and involve comparing similarly manufactured CDs under different environmental cycles. These experiments are nearing their conclusive stage, and the data will be published in 2005.

As the CD manufacturing process continues to be perfected, it is resulting in CDs of different materials with lower and lower BLER levels and improved surface conditions. Even though the observed frequency of failure has been low thus far in our experiments, these improvements will hopefully decrease it appreciably in the future.

Bibliography

1. ANSI X3.199-1991. "Information Systems — 365mm (14.0 inch) Optical Disc Cartridge Write Once — Test Method for Media Characteristics."

2. ANSI AIIM MS59-1996. "Media Error Monitoring and Reporting Techniques for Verification of Stored Data on Optical Digital Data Discs."

3. ANSI AIIM TR39-1996. "Guidelines for the Use of Media Error Monitoring and Reporting Techniques for the Verification of Information Stored on Optical Digital Data Discs."

4. ANSI NPM IT9.21-1996. "Life Expectancy of Compact Discs (CD-ROM) — Method for Estimating, Based on Effects of Temperature and Relative Humidity." Similar reports were produced for magneto-optical (MO) discs and CD-R media.

5. *CD-CATS Users' Manual, SA3*. Audio Development, Inc. www.audiodev.com

6. ISO 9660. "Information Processing — Volume and File Structure of CD-ROM for Information Interchange."

7. ISO/IEC 10149. "Information Technology — Data Interchange on Read-only 120mm Optical Data Disks (CD-ROM)."

Appendix: Summary of error codes

Static data

SLD (start lead) — the optical start of the tracks. This value should be less than 46 mm or the TOC (table of contents) may be missed.

SPD (start of program) — diameter used by the player to locate the start of TRK1 (track 1).

MID (maximum information diameter) — signals the end of the disc. It should be less than 113 mm on a CD-ROM and less than 116 mm on an audio CD. Handling often creates problems closer to the edge. If MID is out of specification, the results may be a weak HF (high frequency) level and disturbed tracking due to birefringence, scratches, and fingerprints from handling.

SVY (scanning velocity) — the speed with which the pits pass the reading spot.

TRP (track pitch) — the distance from the centre of one track to the centre of the adjacent track.

ECC (eccentricity) — the measurement of the geometrical centre of the track of information with respect to the geometrical centre hole of the disc.

Vertical data

DEV (deviation) — specification is ±0.5 mm. Deviation that is too high causes a focussing problem, which in turn causes loss of the HF signal, higher BLER, and possibly the occurrence of E32s.

DEFL (deflection) — the radial deflection of the light beam due to skew. The maximum allowable value in the radial direction is ±1.6°. Deflection that is too high will result in loss of the HF signal.

PP (push pull) — the variation in radial noise. PP is used to translate the wavelength of the noise into an amplitude measurement. It can be measured with either linear polarized light or circular polarized light, and the results will vary depending on the method used.

PPC — PP that is calculated for circular polarized light from linear polarized light according to Philips measurement correlation data.

XT (cross talk) — a measurement of the amount of noise present in the HF signal. XT should be less than 50%. High XT gives a garbled HF signal, which in turn gives higher PP readings.

Dynamic test data

BLER (block error rate) — the number of blocks of data that have at least one occurrence of erroneous data (E11 + E21 + E31). BLER is quantified as the rate of errors per second. BLER is critical because the errors need to be kept to a minimum to ensure data integrity.

E11 — indicates how the existing errors are divided into different severities. Measured in counts per second, E11 indicates the occurrence of single-symbol correctable errors.

E21 — indicates how the existing errors are divided into different severities. Measured in counts per second, E21 indicates the occurrence of double-symbol correctable errors.

E31 — indicates how the existing errors are divided into different severities. Measured in counts per second, E31 indicates the number of first decoder (C1) uncorrectable errors.

E12 — indicates the occurrence of single-symbol correctable errors in the second decoder (C2), measured in counts per second. As E31 errors penetrate to the second decoder, such errors can be further investigated by monitoring the E12, E22, and E32 readings. A high level of E12 is not significant, because one single E31 count can contain up to 30 E12 single-symbol correctable errors, as a result of interleaving.

E22 — indicates a two-symbol correctable error in the second decoder (C2) caused either by one big error or a very high degree of cluster errors. If E22 exists at all, the disc is approaching uncorrectable errors, and there is not much margin of error left to the user.

E32 — indicates uncorrectable errors in the C2 decoder or the existence of unreadable data on the disc. E32 should never be present on a CD.

I3 — the HF level of the shortest pits and land areas of the disc. The I3 level should be kept between 0.3 and 0.7; it should never fall below 0.3 and should preferably be kept above 0.35. Maintaining this level on discs of more than 70 minutes is difficult. I3 levels that are too low result in decoding problems. Scratches and fingerprints from handling give a weak I3 level. I3 is measured by comparing levels with those on a reference disc having known values.

I11 — the HF level of the longest pits and land areas of the disc. The I11 level should be kept above 0.6. Levels that are too low result in decoding problems. I11 is measured by comparing levels with those on a reference disc having known values.

I3R and I11R — the ratios of I3 and I11 to the maximum. I3R and I11R measurements make it possible to distinguish between amplitude variations caused by pit deformation and amplitude variations caused by absorption in the substratum and reflective layer. Measurements are normalized by the reflectivity. I3R and I11R make it possible to see if there are still good pits even though the signal level is low.

REF (reflectivity) — calibrated with a good reference disc and used for detection of reflection variations and normalization of I3R and I11R. Reflectivity is normalized by dividing the maximum level by 0.95.

SYM (symmetry) — given as an absolute value. Specifications for SYM are less than 20% for audio CDs and less than 10% for a CD-ROM. SYM that is too high results in decoder level slicer problems, which would give a BLER reading that is too high.

BERL (burst error length) — a measurement that counts the number of consecutive C2 blocks in error, normally indicating physical damage. Like E22, high BERL leaves little room for error, i.e. handling, dirt, or scratches, after the disc is in use. BERL is important because it indicates the presence of a physical defect large enough to impact more than one block of data.

CRC (cyclic redundancy check) — shows the number of blocks per second where CRC errors are detected. CRC is made on the subcode information.

RN (radial noise) — has a limit of 30 nanometres within the bandwidth of 500–2500 Hz. RN occurs when tracks are damaged. High RN peaks cause the servo to skip tracks. High average RN is an indication of poorly defined pits (worn down stamper or mastering problem).

BEGL (burst errors greater than limits) — with the limit set to 7, this measurement counts the number of occurrences of 7 or more blocks in error.

E32TOT — measurement that counts the total number of uncorrectable blocks present within the testing time. This number should be 0.

Time interval data

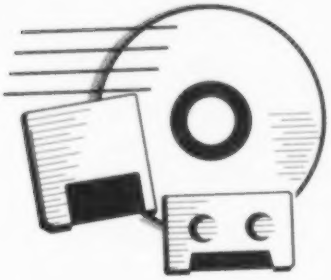
Length deviation — the length of each pit and land area is measured and compared to the ideal pit and land lengths. The result is compared for each run length, 3T to 11T (9 values), with variations preset for each run length.

Jitter — the imprecision in the timing of the discrete samples, pit and land transition, when converting from analog to digital or digital to analog. It involves the tracking, timing of synchronous clock, and any imperfections on the disc. In general this measurement is a good indication of overall disc data quality. Jitter is measured individually for pit and land (3T ...11T), 9 times 2 (18) measurements. Each pit and land is measured and then jitter is calculated statistically as a standard deviation (18 more measurements) according to Philips specifications. This is done for the entire track of the disc.

Résumé

Cette communication porte sur certaines recherches menées à la Bibliothèque du Congrès américain concernant le vieillissement naturel et accéléré des disques compacts sonores. L'étude sur le vieillissement naturel visait à suivre les effets à long terme de conditions normales d'entreposage et d'utilisation sur l'intégrité d'un échantillon restreint de disques compacts sonores choisis au hasard dans la collection de la Bibliothèque. Au cours de cette recherche, les disques compacts de l'échantillon sont restés dans la collection et ont continué de faire l'objet d'une usure normale; ils ont été retirés très brièvement de la collection tous les deux ou trois ans pour effectuer des essais. Les résultats obtenus au cours des sept dernières années de ce projet pilote sont exposés, ainsi que leurs incidences sur son orientation future. On présente également l'état d'avancement d'une étude en cours sur le vieillissement accéléré des disques compacts sonores, ainsi que diverses observations visuelles concernant des défauts entraînés par le processus de vieillissement accéléré.

L'objectif de ces études est de comprendre l'importance de tout ce qu'il faudra probablement faire pour préserver à long terme une collection contemporaine de disques compacts qui doit inévitablement comprendre une série de caractéristiques de composition à chaque étape de sa fabrication. Nous espérons que ces recherches nous permettront éventuellement : 1) de protéger le mieux possible à long terme nos collections; 2) d'établir des paramètres pertinents de vérification des disques compacts qui permettraient de déterminer le moment où il faut faire des copies de secours ou faire migrer les données; 3) de reconnaître les problèmes qui résultent des défauts de fabrication des disques compacts.



RESTLESS RECORDS: THE STATE OF SMALL VIDEO ART ARCHIVES IN CANADA

MAIJA MARTIN

#6-1336 East 2nd Avenue
Vancouver BC V5N 1C5
Canada
Tel.: (604) 215-4110
E-mail: maijamartin@telus.net

Abstract

This paper provides a survey of the efforts of a number of small artist-run media art centres across Canada to preserve their videotape collections. Included in the discussion is an account of some of the technical approaches used by these centres, such as cleaning and re-mastering methods, preferred formats for storage, and proper storage environments. Challenges identified by these centres include lack of funding and expertise, obsolescence of playback equipment, and a lack of standards for preservation.

Also included in this paper is a draft guide to videotape preservation for small video archives in Canada. This guide was prepared in an attempt to create standards for smaller videotape archives with regards to technical approaches and storage options.

Introduction

This paper provides an overview of the efforts of small artist-run media art centres across Canada to preserve their videotape collections. It is my opinion that the lack of access to information and insufficient funding have prevented the establishment of achievable standards for smaller independent media art archives.

My goal in writing this paper is to build the skeleton of a manual for smaller media art archives in Canada, to be ratified by the various archives. With the input of conference participants, I would like to establish sustainable standards for these centres.

Survey

For the purposes of this paper, a small archive is defined as an archive that typically has a collection of under several thousand titles and receives little or no funding for its archival efforts. For most of these organizations (e.g. video distribution centres and small art galleries), their primary mandate does not include archival responsibilities. While following the models of larger archives such as the National Film Board and Library and Archives Canada would be extremely useful for these smaller archives in terms of identifying standards, many of them lack the staff, expertise, and funding to be able to carry out the preservation strategies laid out by larger archives. For these organizations much of the video restoration expertise has historically been concentrated in one staff member, which is especially problematic in centres with high staff turnover.

The centres surveyed for this paper include Groupe Intervention Vidéo (Montreal), Video Out Distribution (Vancouver), Video Pool (Winnipeg), Vidéographe (Montreal), VTape (Toronto), and Western Front Media (Vancouver). Each was questioned about the size of its collection, previous restoration projects, and obstacles to carrying out a preservation plan.

The main challenge faced by these centres is the obsolescence of the tape formats and the playback machinery needed to view the material. Since many of the tapes have not been played in years, they have become "sticky" and need to be cleaned before they can be played now.

Different centres chose different cleaning methods and in some cases different tape formats for re-mastering. In all cases, the centres were unable to re-master older works onto ideal tape formats due to lack of funds. That is to say, while it is thought that the DigiBeta format may be the best preservation format for the moment (although there are murmurings of its obsolescence), none of the centres has had access to the funding that would be required to transfer its entire collection to this format.

Overview of the centres surveyed and their preservation histories

Western Front Media (Vancouver, BC)
1000 titles
Western Front Media undertook a re-mastering project (funded by the Daniel Langlois Foundation) in 2001 to restore its 1/2-inch open-reel and 3/4-inch tapes. The tapes were prepared using the baking method (see the guide in the Appendix), cleaned using a modified open-reel machine, transferred to 3/4-inch SP tape, and digitized onto DVD for viewing purposes.

Video Out Distribution (Vancouver, BC)
2500 titles from 1972
Video Out Distribution has undertaken several restoration projects over the years. In the Legacy Project in 1992, tapes on 1/2-inch open-reel were prepared using the baking method, cleaned using a modified open-reel machine, and transferred to 3/4-inch Sony tapes. More recently Video Out Distribution commissioned a report on how to proceed with the remaining open-reel and now rapidly deteriorating 3/4-inch tapes.

Vidéographe (Montreal, QC)
1100 titles; oldest tape 1970
Vidéographe undertook a restoration project in 1998, using a machine designed to remove the sticky outer layer of 1/2-inch open-reel tapes. The tapes were then transferred to BetaSP, and the masters are now stored in the vault at the Cinémathèque Québécoise.

Video Pool (Winnipeg, MB)
200 titles
In 2000, Video Pool constructed a "vault" (consisting of a room within a room and an air conditioner to control temperature and humidity) in which to store its masters.

VTape (Toronto, ON)
2500 titles
VTape actively collects older videotape archives, and provides a transfer service from 1/2-inch open-reel and 3/4-inch to more stable formats.

Groupe Intervention Vidéo (Montreal, QC)
300 titles
Groupe Intervention Vidéo currently stores its masters on 3/4-inch and miniDV, and is currently upgrading its submasters to miniDV.

Breakdown of preparation and cleaning methods used by the centres

The centres vary in terms of methods used to clean older videotapes. For details on the preparation and cleaning methods, as well as alternative methods, please see the guide in the Appendix.

1. Baking
Tape baking is a controversial process in which tapes are heated at a low temperature for 5–6 hours in order to allow for a cleaner, less sticky playback.

Although the baking process has not been discounted by all archivists, many of the centres surveyed did not approve of this method because of the dangers involved. However it is a method that has been used by a number of centres with much success, including Western Front Media and, historically, Video Out Distribution. Several members of the AMIA (Association of Moving Image Archivists) have also used the method with success.

2. High-speed cleaning
Vidéographe and VTape have both constructed machines for cleaning older videotapes that remove the sticky outer layer. Although both of these centres had great success with these machines, other centres found that the method was not sufficient for removing the dirt from the tapes. Some centres had more success if tapes were baked prior to cleaning.

3. Professional out-of-house cleaning
Many of the organizations surveyed did not have access to the funding that would be required to send entire collections of tapes out-of-house for cleaning, but reserved it as an option for problematic tapes.

Preferred tape formats

All of the centres identified BetaSP as their preferred storage format. DigiBeta was beyond their financial reach,[1] and 3/4-inch formats were considered outdated.

Storage areas

Only the National Film Board in Montreal and Library and Archives Canada in Ottawa have proper temperature-controlled storage vaults. Most centres store their masters in less than ideal situations (i.e. non-humidity-controlled rooms), with the exception of Vidéographe and Video Pool. The other centres do their best to store the tapes in clean rooms with as little temperature fluctuation as possible.

Sources of funding

Only two of the organizations received funding specifically for an archive project: Western Front Media (from Daniel Langlois Foundation, now a discontinued program) and Video Pool (from a municipal program called Centre Venture and private donations).

Conclusion

Despite obstacles such as lack of funding and access to information, smaller media art archives in Canada have been working towards preserving the older videotapes in their collections. In order to aid them in their efforts, some kind of communication tool such as a listserv should be established through which they can share information about possible sources of funding, cleaning methods, and storage formats. A manual should be created as a starting point for establishing a dialogue around these issues. Such dialogue could result in concrete, lower-budget options for centres that cannot afford professional out-of-house cleaning.

Appendix: Guide to videotape preservation for small archives

Videotape is not considered an archival format, but many internationally renowned Canadian artists have chosen to work in the medium. Such works are often stored in independent distribution centres, galleries, and artist-run centres across the country. Unfortunately, many of the videotape formats on which these works were created and stored are becoming obsolete, and the works are in danger of being lost forever.

The following is a guide to assessing, cleaning, and re-mastering these older videotapes. It is intended for artist-run centres that are interested in preserving their videotape collections but are unfamiliar with formal archival processes.

Damage assessment and prioritizing

Challenges to restoring older videotapes fall into three main areas:

- obsolete formats and playback equipment
- sticky residue or powder on tape and binder degradation (oxide flaking off the base film)
- disasters such as tape erasure, exposure to magnetic fields, flooding, or fire

Before beginning the inspection process, it is important to determine which tapes in the collection should be restored and in what order (i.e. prioritize the collection). Factors influencing this decision should include:

- the value of the tape, as determined by its uniqueness or scarcity and its relationship to the organization
- the rate of deterioration of the tape, determined by its age, format, rate of extinction (see Table 1), and how it has been stored
- legal ownership of the tape
- the tape's proximity to the original, whether it is a master, submaster, or exhibition copy

Table 1. Obsolescence ratings of some common tapes
(compiled by Paul Messier, Boston Art Conservation)

2" Quadruplex or Quad, Ampex Corporation (1956)	Critically endangered
TCR-100 "Quad" Video Cartridge, RCA (1969)	Critically endangered
1" UV-340 or EV 210, Sony (1964)	Extinct
1" SMPTE Type A, Ampex Corporation (1965): 1" 7100 (Precision Instruments) 1" IVC 800/900 1" Helical 1" SMPTE Type B (introduced by Bosch, 1975) 1" SMPTE Type C (1978)	Critically endangered or Endangered
EVR Electronic Video Recording, CBS (1967)	Extinct
Panasonic Video Cartridge	Extinct
EIAJ Type 1, Sony Broadcasting and others (1969)	Endangered
Philips Videocassette Recorder, N.V. Philips (1970)	Extinct
U-Matic, Sony, JVC, Panasonic and others (1971)	Threatened
ANSI/SMPTE 21M-1997	
Cartrivision, AVCO Corporation (1972)	Extinct
V-Cord, Sanyo (1972)	Extinct
BetaMax, Sony (1975)	Critically endangered
VHS, JVC/Panasonic (1976–77)	Lower risk (however, if the tape itself is older than 10 years, it should be considered for recopying)
Betacam SP, Sony (1986)	Threatened
M-II, Panasonic/JVC (1986)	Vulnerable
S-VHS, JVC and others (1987)	Vulnerable
D-2, Ampex and others (1988)	Threatened
Hi8, Sony and others (1989)	Threatened
D-3, Panasonic (1990)	Lower risk
Digital Betacam, Sony (1993)	Vulnerable
Digital-S, JVC (1995)	Vulnerable

Tape inspection techniques[2]

The following inspection procedures should be performed in the order listed. If any sign of fungus is detected in Step 2, inspection should be stopped immediately.

1. Check the physical container for damage to its structural integrity. Breakage is a strong indication of improper handling. If the container, reel, or cassette is damaged, the tape inside is also likely to have suffered damage or contamination.

2. Check the interior of the container and the edges of the tape for patterned black-, brown-, or mustard-coloured contamination and for fuzzy or thread-like growths that indicate the presence of fungus. Fungus can grow on tape after it has been exposed to high humidity. Tapes with fungus should be isolated and treated by professionals as soon as possible. *Tape with fungus can present a health hazard: if fungus is suspected do not continue the inspection!*

3. Smell the tape as soon as it is removed from its container. Hydrolytic breakdown of polyester binder creates esters that have distinctive odours but dissipate quickly. The most common odours can be characterized as "waxy," "dirty socks," or "astringent/pungent" depending on the binder. Some early tapes using an acetate base will also give off an odour of "vinegar" if the base is beginning to decay. Tapes with binder hydrolysis are in the process of self-destruction and can stick in the machine during playback, causing additional damage. Binder hydrolysis can be treated, but acetate breakdown is permanent.

4. With a light source above and slightly behind, tilt the tape edge-on at approximately 45° away from the light source and inspect the tape pack. Check for spoking, popped strands, stepped pack, edge damage, "shiners," and windows. These terms refer to irregularities in the way the tape is wound onto the hub and are indications of improper handling, storage, or a badly set up machine. Spoking shows up as a pattern radiating out from the hub and is the result of improper tension. Popped strands and stepped pack refer to individual or groups of tape wraps that stick up from the edge of the pack. When light is shone on the tape, "shiners" show up as thin strips of greater reflection resulting from the edge of the tape being torn or folded. Windows are gaps in the tape pack caused by the tape becoming loose on the hub and often indicate a place where the tape has folded back over on itself.

5. Check the tape edge and the reel/cassette/cartridge for particulate contamination and for signs of staining that may indicate liquid contamination. Any visible contamination is an indication of poor storage or handling. Particulate contamination can block the signal during playback and can scratch both the tape and the playback heads. Liquid contamination will accelerate tape decay and can often result in tape wraps sticking together.

6. Check the tape edge for white powder or crystalline residue and check the interior of the container for black/brown flakes of oxide. These symptoms, caused by a variety of conditions, indicate that the tape is beginning to break down.

7. If the tape is reel-to-reel, allow a few outer wraps to hang loose and examine for physical distortion and binder/base adhesion failure. Physical distortion caused by improper tension will often show up on a loose piece of tape as wavy or "scalloped" edges. Binder/base failure is identified by sections of the tape that appear to be a different colour where the binder has come off. Note: If transparent or masking tape has been used to secure the tape end, binder loss in the isolated area under the adhesive is not an indication of binder/base failure.

Preparation and cleaning options

There are several preparation and cleaning options available for sticky tapes.

1. Baking method
The baking method is used to treat tapes that suffer from sticky tape syndrome (i.e. tapes in which the binder is flaking off and creating a gummy residue). The process involves heating the tape at a very low temperature for a period of 5–6 hours, and then cooling it for an equal amount of time. This results in a partial re-absorption of oligomer residue and the cross-linking of molecules into long-chain polymers, allowing for a cleaner playback.

The biggest danger in the baking process is that the temperature necessary to create cross-linking is very close to the temperature at which the tape particles fall out of their recorded alignment, erasing the video signal. Others dangers involve the baking of tapes that have not been rewound before they are stored, and possible damage to a tape if it is baked repeatedly (particularly dangerous if the tape is an original).

While the baking process has not been discounted by all archivists, many do not approve of it because of the dangers involved. However, a number of centres have used the method with some success.

2. High-speed cleaning
This cleaning option is carried out by fast-forwarding and rewinding videotapes while passing the outer edge of the tape across a lint-free cloth until all dirt has been removed. A machine to accomplish this can be constructed out of an open-reel machine in which the video heads are beyond repair. This is the safest method and exposes the tapes to the least potential damage, but it may not be sufficient to clean excessively dirty tapes.

3. Storage method
An alternative to baking or high-speed cleaning is to place the tapes in a stable environment with a moderate to low temperature and a relative humidity that is below 30%, and leave them there for a year or two. This will also cause cross-linking and reduce or eliminate the sticky effect of hydrolysis.

4. Fridge method
In this case the tapes are refrigerated in double freezer bags with molecular sieves (available from Kodak) for a period of 1 week, and then copied to a more stable format. Some archivists recommend a longer period of refrigeration — between 5 and 6 weeks. Note that the tapes still need to be cleaned after they have been stabilized.

5. Professional out-of-house cleaning
If it is not possible to clean a tape using in-house methods, the following organizations/companies provide professional cleaning services:

- Bay Area Video Coalition
- VTape
- Vidipax
- Specs Brothers

Storage formats

Several factors should be considered when choosing a format for re-mastered tapes:

- longevity of the tape format and its playback equipment
- generation loss
- cost

DigiBeta is currently the storage format preferred by archivists, but it may not be an option for many centres due to its cost. BetaSP provides a possible alternative as it is a broadcast standard format and the accompanying equipment is likely to be available on the market for some time.[3]

DVDs are a good exhibition format, and may save on storage space, but the level of compression required is not currently suitable for archival purposes. Used tapes are also not recommended for storage purposes.

Storage area

The main requirement for a storage area is that it be dry and cool — the colder and drier the environment, the better for the tapes. However, this type of storage can make it difficult to access tapes because they must be returned to room temperature in a controlled environment (which requires a fair bit of time and special handling procedures) before they can be safely used. Also, it is important to avoid temperatures that are too low. In some cases, temperatures lower than 32°F may actually harm the media and shorten life expectancies by risking exudation of the lubricant from the binder, which may clog the heads.

Basic maintenance tips

1. Rewind tapes before storing so that if there is damage it is over colourbars and tone, not content.
2. Fast-forward and rewind tapes in playback (shuttle) mode.
3. Store cassettes on their edges so that water will not be trapped in the event of a flood.

Disaster plan

For disaster plan suggestions, see the Web site of Specs Brothers (www.specsbros.com/disaster.html).

Sources of funding

The following is a list of Canadian funding bodies that can be approached for media art preservation projects. Other options open to centres include raising funds through events such as screenings, as well as soliciting private donations.

Canadian Council of Archives (CCA)
344 Wellington Street, Room 1009
Ottawa ON K1A 0N3
Tel.: 1-866-254-1403
Fax: (613) 947-6662
E-mail: nlavigne@archives.ca

The CCA has a number of funding programs including the Control of Holdings Program through which centres can receive assistance for the processing of audiovisual material. To be eligible for such assistance, a centre must be a member of the provincial branch of the CCA, and must have its archival material sorted according to the Rules for Archival Description.

Young Canada Works in Heritage Institutions (YCW)
c/o The Heritage Canada Foundation
5 Blackburn Avenue
Ottawa ON K1N 8A2
Tel.: (613) 237-1066
Fax: (613) 237-5987
E-mail: youngcanadaworks@heritagecanada.org

YCW provides funding to hire youth to work in heritage organizations.

Canada's Digital Collections (CDC)
Industry Canada
Jean Edmonds Tower
300 Slater Street, 18th Floor, Section D
Ottawa ON K1A 0C8
Tel.: 1-800-465-7766
Fax: (613) 941-1232
E-mail: collections@ic.gc.ca

CDC offers funding for multimedia projects that provide work experience for young Canadians aged 15–30. These projects should involve digital conversion of analog material and the creation of Web sites for its display online.

Samuel and Saydie Bronfman Family Foundation (SSBF)
1170 Peel Street, Suite 800
Montreal QC H3B 4P2
Tel.: (514) 878-5270
Fax: (514) 878-5293
E-mail: info@bronfmanfoundation.org
www.bronfmanfoundation.org

The SSBF provides matching funding up to $5000 through its Special Initiatives Program for new or established organizations wishing to undertake a new initiative within their existing program of work.

Further research

The following sources offer more information about preserving videotapes.

Independent Media Arts Preservation
www.imappreserve.org

Specs Bros.
www.specsbros.com

Extinct Videotape Recorder links
www.labguysworld.com/VTR_Links.htm

Society of Motion Picture and Television Engineers (SMPTE)
www.smpte.org

Vidipax
www.vidipax.com

Estate Project
www.artistswithaids.org/

American National Standards Institute (ANSI)
www.ansi.org

Conservation On Line (CoOL)
palimpsest.stanford.edu/

Library of Congress
(TV/Videotape Study for Congress)
lcweb.loc.gov/film

National Media Lab/National Technology Alliance
www.nta.org/MediaStability

Video Preservation Basics
www.experimentaltvcenter.org/history/preservation/preservation.php3

Television/Video Preservation Study: Volume 1: Report
lcweb.loc.gov/film/tvstudy.html

Fading Away: Strategic Options to Ensure the Protection of and Access to Our Audio-Visual Memory. Report of the Task Force on the Preservation and Enhanced Use of Canada's Audio-Visual Heritage. Ottawa, ON: National Archives of Canada, June 1995, 51 pp.

Playback: A Primer for Video Preservation.
Bay Area Video Coalition, 1997.

Preserving Our Heritage. Keith Donohoe,
Center for Arts and Culture, 2001.

Endnotes

1. A 60-minute BetaSP tape costs around CAN$40, compared to CAN$70 for a DigiBeta tape.

2. Specs Bros. *Basic Inspection Techniques to Sample the Condition of Magnetic Tape.* 2002. www.specsbros.com/whitepaper.html

3. It should, however, be noted that Sony has stopped the production of two of its three BetaSP lines, and the usefulness of BetaSP as a storage medium will need to be re-assessed in the upcoming years.

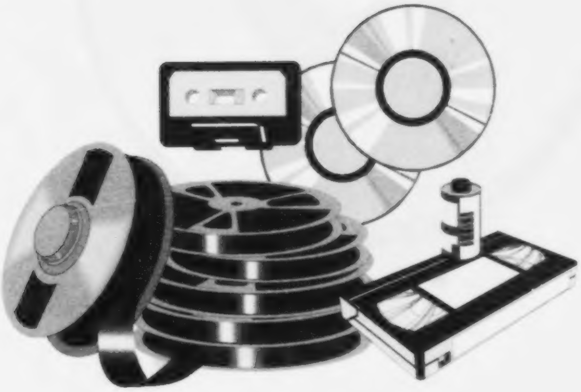
Résumé

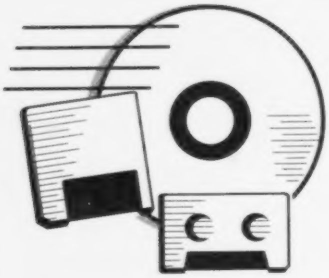
Nous parlerons des efforts de petits centres d'art canadiens gérés par des artistes pour préserver leurs collections de bandes vidéo. Nous rendrons notamment compte des techniques employées par ces centres, par exemple le nettoyage et la remasterisation, les modes d'entreposage adoptés et les environnements propres à l'entreposage. Les difficultés circonscrites par ces centres sont les suivantes : manque de fonds et de savoir spécialisé, obsolescence du matériel de lecture, et manque de normes de préservation.

Nous offrirons aussi une ébauche de guide pour la préservation des bandes vidéo dans les petits services d'archivage magnétoscopique du Canada. Ce guide a été rédigé dans le but d'essayer de créer des normes concernant les techniques utiles et les modes d'entreposage possibles pour ces petites archives.

POSTERS

AFFICHES





Survey Results: Preservation Management of Born-digital Documents in Manuscript Repositories in the United States

Ruth Elizabeth Bryan

Rare Book, Manuscript, and Special Collections Library
Box 90185, Perkins Library
Duke University
Durham NC 27708-0185
USA
Tel.: (919) 660-5982
E-mail: ruth.bryan@duke.edu

Purpose

Manuscript repositories that collect papers from private individuals and records from outside organizations are usually unable to communicate with the creators of these electronic documents prior to donation to ensure appropriate preservation practices. A survey was therefore conducted to determine how well developed the preservation programs were at selected manuscript repositories in the United States.

Method

The survey was conducted in April 2003 through telephone interviews with nine manuscript repositories in the United States.

Principal results

1. Four of the repositories surveyed print electronic documents to paper, and three migrate them to a server. Of these three, one has developed an untested patron access procedure. The predicator for printing versus migrating was the type of software received.
2. Five of the repositories converse with donors regarding preservation.

Conclusion

Although basic preservation measures for electronic records have been established, most of the repositories that were surveyed feel they do not yet know enough to develop a full preservation program. Workflow considerations (the "who" and "when") are as important as the "how" but are less discussed. Because manuscript material is private, authenticity requirements may be looser than those for public material.

Objet

Les services d'archivage de manuscrits qui recueillent les documents de particuliers et les registres d'organismes extérieurs ne peuvent généralement pas communiquer avec les auteurs de documents électroniques pour garantir la qualité des pratiques de conservation avant la donation. Une étude a donc été menée pour savoir où en sont les programmes de préservation de certains services américains d'archivage de manuscrits.

Méthode

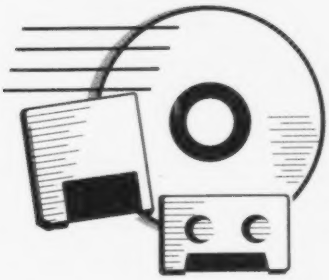
Enquête téléphonique auprès de neuf services américains d'archivage de manuscrits en avril 2003.

Principaux résultats

1. *Quatre services d'archivage impriment les documents électroniques, et trois les acheminent vers un serveur. Un de ces trois services a élaboré une procédure d'accès non vérifiée pour les usagers. C'est le type de logiciel reçu qui détermine le choix entre impression sur papier et migration vers un serveur.*
2. *Cinq services d'archivage communiquent avec les donateurs concernant les méthodes de préservation.*

Conclusion

Il existe des mesures de préservation de base pour les documents électroniques, mais la plupart des services d'archivage interrogés ont le sentiment qu'ils n'en savent pas assez pour élaborer un programme complet de préservation. Les considérations relatives à l'acheminement du travail (qui et quand) sont tout aussi importantes, mais on en discute moins que des méthodes (comment). Comme le contenu des manuscrits est privé et non public, les exigences en matière d'authenticité sont peut-être moins strictes.



VIRTUAL REMOTE CONTROL: PRESERVATION RISK MANAGEMENT FOR WEB RESOURCES

NANCY Y. MCGOVERN, ANNE R. KENNEY, RICHARD ENTLICH,
WILLIAM KEHOE, AND PETER BOTTICELLI

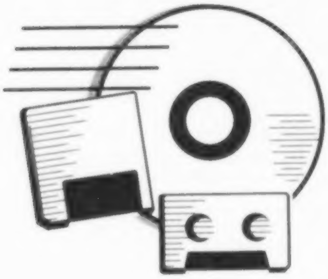
Cornell University
215 Olin Library
Ithaca NY 14850
USA
Tel.: (607) 255-1436
E-mail: nm84@cornell.edu

Abstract

The rapid growth of the World Wide Web provides extensive Web sites and pages that may be neither owned nor controlled by an organization, but may provide essential enhancements to existing resources or form the nucleus for new collections or series. The Virtual Remote Control (VRC) approach adapts the risk management model (risk identification, classification, assessment, analysis, and program implementation) to the monitoring and evaluation of changes to ensure the longevity of the Web. VRC consists of six stages: Identification, Analysis, Appraisal, Strategy, Detection, and Response. Each stage has an organization and a toll-based component that work in combination. The approach may be used by an organization to manage its own Web resources or by a digital preservation management program to manage targeted Web resources over time. It will be used to predict probable loss based on the presence or absence of key indicators that may enable or inhibit longevity.

Résumé

La croissance rapide de la Toile donne accès à de vastes sites ou pages Web qui n'appartiennent à aucune organisation ou ne sont contrôlées par aucune organisation, mais qui peuvent comporter des éléments d'amélioration indispensables aux ressources actuelles ou constituer le noyau de nouvelles collections ou séries. La méthode du contrôle à distance virtuel (VRC) est une adaptation du modèle de gestion du risque (circonscrire, classer, évaluer et analyser les risques, puis mettre le programme en œuvre) au suivi et à l'évaluation des changements pour garantir la longévité de la Toile. Le contrôle à distance virtuel comprend six étapes : circonscrire, analyser, évaluer, élaborer une stratégie, dépister et réagir. Chaque étape comporte un élément organisationnel et un élément instrumental qui fonctionnent de concert. La méthode peut être employée par une organisation pour gérer ses propres ressources sur Internet ou dans le cadre d'un programme de gestion de la préservation numérique pour gérer la durée de vie de ressources ciblées sur Internet. Elle servira à prévoir les pertes probables en fonction de la présence ou de l'absence de certains indicateurs favorisant ou inhibant la longévité.



ROLE OF LACQUER AND INK IN PROTECTING COMPACT DISCS EXPOSED TO GASEOUS POLLUTANTS

LÉON-BAVI VILMONT

Centre de recherches sur la conservation des documents graphiques (CRCDG)
36 Geoffroy Saint-Hilaire
75005 Paris
France
Tel.: (33) 01 44 08 69 98
E-mail: vilmont@mnhn.fr

Abstract

A preliminary study has shown that NO_2 and SO_2 gases cause degradation of the reflective layer of compact discs (starting at the outer and inner edges) but do not appear to affect the parts covered with lacquer and ink.

To determine the extent to which lacquer and ink act as a protective layer, we had samples of discs specially manufactured with an aluminum layer, some with a protective lacquer coating and others without, and some covered with ink and others not. The samples were exposed to 10 ppm of NO_2 and 5 ppm of SO_2 at 23°C and 50% RH, either combined or separately. The data error rates generated by the discs were regularly measured using a CD analyser.

Exposure of the CDs to gaseous pollutants resulted in a reduced data recovery. The rate of degradation was much faster with NO_2 alone than with NO_2 mixed with SO_2. The application of ink adversely affected the discs. The protective role of lacquer was confirmed.

Résumé

Une étude préliminaire avait montré que les gaz NO_2 et SO_2 provoquaient une disparition de la couche réfléchissante des disques compacts, à partir des bords extérieur et intérieur, mais semblait épargner les parties recouvertes de vernis et d'encre.

Pour savoir dans quelle mesure le vernis et l'encre pouvaient servir d'écran de protection, nous avons fait fabriquer spécialement des échantillons de disques avec une couche métallique d'aluminium pourvu ou non de vernis protecteur, recouvert ou non d'encre. Les échantillons sont exposés à 10 ppm de NO_2 et 5 ppm de SO_2 à 23 °C et 50 % HR, pris en mélange ou séparément. Les taux d'erreurs générées par les disques sont mesurés régulièrement au moyen d'un analyseur de disques compacts.

L'exposition aux gaz polluants conduit à une perte de la qualité de restitution des informations des disques. La dégradation est beaucoup plus rapide avec NO_2 seul qu'en mélange avec SO_2. L'application d'encre a un effet néfaste sur les disques. En revanche, le rôle protecteur du vernis est confirmé.

ISBN 0-662-68620-9
9 780662 686200